Cloning and characterization of two glycosidases from the acidothermophile *Alicyclobacillus acidocaldarius* ATCC27009

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium
(dr. rer. nat.)
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Herr Dipl.-Biol. Kelvin Eckert
geboren am 18.12.1971 in Berlin

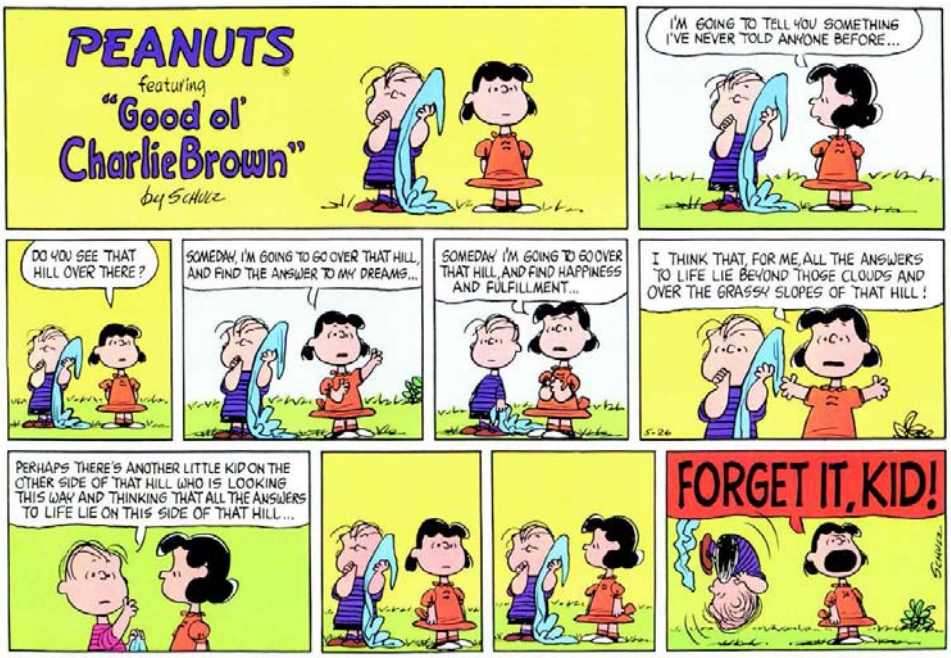
Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Michael Linscheid

Gutachter:

1. Prof. Dr. rer. nat. Erwin Schneider
2. Prof. Dr. rer. nat. Rainer Borriss
3. Prof. Dr. rer. nat. Evert Bakker

eingereicht am: 17. September 2003
Tag der mündlichen Prüfung: 6. Februar 2004


PEANUTS
featuring
"Good ol' CharlieBrown"
by Schulz
I'M GOING TO TELL YOU SOMETHING I'VE NEVER TOLD ANYONE BEFORE...
DO YOU SEE THAT HILL OVER THERE?
SOMEDAY, I'M GOING TO GO OVER THAT HILL, AND FIND THE ANSWER TO MY DREAMS...
SOMEDAY I'M GOING TO GO OVER THAT HILL, AND FIND HAPPINESS AND FULFILLMENT...
I THINK THAT, FOR ME, ALL THE ANSWERS TO LIFE LIE BEYOND THOSE CLOUDS AND OVER THE GRASSY SLOPES OF THAT HILL!
5-26
PERHAPS THERE'S ANOTHER LITTLE KID ON THE OTHER SIDE OF THAT HILL WHO IS LOOKING THIS WAY AND THINKING THAT ALL THE ANSWERS TO LIFE LIE ON THIS SIDE OF THAT HILL...
FORGET IT, KID!
SCHULZ

Part of this work has been published in the following papers:

Eckert, K. and Schneider, E. (2003). A thermoacidophilic endoglucanase (CelB) from *Alicyclobacillus acidocaldarius* displays a high sequence similarity to arabinofuranosidases belonging to family 51 of glycoside hydrolases. *Eur. J. Biochem.* **270**: 3593-3602.

Eckert, K., Ernst, H.A., Schneider, E., Larsen, S. and Lo Leggio, L. (2003). Crystallization and preliminary X-ray analysis of *Alicyclobacillus acidocaldarius* endoglucanase CelA. *Acta Crystallogr. D. Biol. Crystallogr.* **59**: 139-141.

Eckert, K., Zielinski, F., Lo Leggio, L. and Schneider, E. (2002). Gene cloning, sequencing, and characterization of a familiy 9 endoglucanase (CelA) with an unusual pattern of activity from the thermoacidophile *Alicyclobacillus acidocaldarius* ATCC27009. *Appl. Microbiol. Biotechnol.* **60**: 428-436.

Contents

1 Abbreviations

aa	amino acids
ABC	ATP-binding-cassette
C	cellobiose
CBD	cellulose binding domain
CBM	carbohydrate binding module
$CelB_{trunc}$	heterologous truncated CelB protein
CMC	carboxy methyl cellulose
CRE	catabolite responsive element
Dig	digoxigenin
DP	degree of polymerization
DTT	1,4-dithio-DL-threitol
EC	Enzyme Commission number
EDTA	ethylenediaminetetraacetate
EDXRF	energy dispersive X-ray fluorescence spectrometry
fig.	figure
G	glucose
G3	cellotriose
G4	cellotetraose
GH	glycoside hydrolase
Hepes	*N*-2-Hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
Ig	immunoglobulin
IUB-MB	International Union of Biochemistry and Molecular Biology
IPTG	isopropyl β-D-thiogalactoside
kb	kilo base pairs
LB	Luria-Bertani medium
MCS	multiple cloning site
MPD	2-methyl-2,4-pentanediol
nt	nucleotides
ORF	open reading frame
pNP	*p*-nitrophenol
pNParaf	*p*-nitrophenyl-α-L-arabino-furanoside
pNPC	*p*-nitrophenyl-β-D-cellobioside
pNPG	*p*-nitrophenyl-β-D-glucoside

pNPG3	*p*-nitrophenyl-β-D-cellotrioside
pNPG4	*p*-nitrophenyl-β-D-cellotetraoside
rpm	rounds per minute
RT	room temperature
sec.	section
SLH	S-layer homology
tab.	table
TLC	thin layer chromatography
Tris	tris(hydroxymethyl)aminomethane
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

2 Summary

2.1 English

Two glycoside hydrolases from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* were characterized and the corresponding genes cloned and sequenced. Together the intracellular enzyme CelA and the extracellular enzyme CelB may play a major role in the degradation of β-1,4-linked polysaccharides.

A gene encoding a β-1,4-endoglucanase (CelA) was cloned and the enzyme overexpressed and purified. The protein contained an immunoglobulin-like domain but lacked a cellulose-binding domain. In conjunction with sequence similarities of the catalytic domain these features demonsrated the protein to be a member of glycoside hydrolase family 9, subgroup E1.

CelA was most active against substrates containing β-1,4-linked glucans, but also exhibited activity against oat spelt xylan. It displayed a pH optimum of 5.5 and a temperature optimum of 70 C. The protein was found to contain one zinc and two calcium ions, likely to be important for temperature stability. It showed a striking pattern of hydrolysis on p-nitrophenyl glycosides, with highest activity on the cellobioside derivative, some on the cellotetraoside derivative and none on the glucoside and trioside derivatives. For chromogenic activity against p-nitrophenyl cellobioside a K_{m} of 3.0 mM and a k_{cat} of 7.1 s^{-1} (V_{max} of 7.13 U mg^{-1}) were determined. The hydrolysis patterns led to the conclusion that CelA contained a steric block for β-1,4-linked glucans on the non reducing side of subsite -2 or, alternatively, two strong binding sites -1 and -2. No signal peptide for transport of CelA across the membrane was detected. This, together with high activity on oligosaccharides, a near neutral pH optimum, and inactivation at low pH, suggests a role for the protein as a cytoplasmic enzyme for the degradation of imported oligosaccharides.

A second enzyme with xylanase and cellulase activity was purified from *A. acidocaldarius* cultures. CelB displayed a molecular mass of 100 kDa and could only be removed from cells with the help of detergent. Cloning and sequence analysis of the corresponding gene revealed an ORF encoding a preprotein with a typical sec-dependent signal peptide. Purified recombinant

CelB and a truncated varient lacking the C-terminal 203 amino acid residues displayed enzymatic properties similar to the wild-type protein. A low pH optimum of 4 was found. Stability was also high at low pH, the enzyme retaining 80 % of activity after incubation over night from pH 1.5 to 7. The temperature optimum was 80 C, a temperature at which the enzyme was also stable, showing 60 % residual activity after 1 h. CelB displayed an endo mode of action, but release of cellobiose and xylobiose from cellulose and xylan, respectively, was observed after prelonged periods of incubation. CelB belonged to glycoside hydrolase family 51, but it was only the second entry in this family with activity typical of an endoglucanase. Highest sequence similarity was found towards the other endoglucanase, EGF from *Fibrobacter succinogenes*, the two forming a distinct group in the phylogenetic tree of this family.

Analysis of the amino acid composition of the catalytic domains demonstrated that CelB contains fewer charged amino acids than its neutrophilic counterparts, which is in line with adaptation to low pH.

Wild-type and full-length recombinant CelB were soluble only in detergent. In contrast, truncated CelB was completely water soluble, suggesting a role of the C-terminal region in cell association. This C-terminal hydrophobic region displayed local sequence similarities to a hydrophobic region of an amylase from the same organism.

2.2 German

Zwei Glykosylhydrolasen des acidothermophilen Bakteriums *Alicyclobacillus acidocaldarius* konnten charakterisiert werden. Die jeweiligen Gene wurden kloniert und sequenziert. Das intrazellulre Enzym CelA und das extrazellulre Enzym CelB knnten zusammen eine tragende Rolle im Abbau von β-1,4-verknpften Polysacchariden spielen.

Ein Gen, welches fr eine β-1,4-Endoglucanase (CelA) kodiert, wurde kloniert und berexpremiert. Das Enzym wurde gereinigt. Das Protein besitzt eine Immunoglobulin-hnliche Domne, jedoch keine Cellulosebindedomne. Zusammen mit den Sequenzhnlichkeiten der katalytischen Domne zeigen diese Merkmale, da das Protein zur Familie 9, Untergruppe E1 der Glykosylhydrolasen gehrt.

CelA zeigte hchste Aktivitt gegenber β-1,4-Glucanen, besa jedoch auch Aktivitt gegen Haferspelzenxylan. Das Enzym hatte ein pH optimum von 5.5 und ein Temperaturoptimum von 70 C. Im Protein waren zwei Zink- und zwei Calcium-Ionen gebunden, die wahrscheinlich wichtig fr die Temperaturstabilitt sind. Gegenber *p*-Nitrophenylglykosiden ergab sich ein berraschendes Hydrolysemuster: Hchste Aktivitt wurde auf dem Cellobiosidderivat gefunden, eine niedrigere Aktivitt fand sich auf dem Cellotetraosederivat, wohingegen keine Aktivitt auf den Glucose- und Cellotriosederivaten gemes-

sen wurden. Fr *p*-Nitrophenylcellobiosid wurde ein K_{m} von 3.0 mM und ein k_{cat} von 7.1 s^{-1} (V_{max} = 7.13 U mg^{-1}) in Bezug auf chromogene Aktivitt bestimmt. Das Hydrolysemuster fhrte zu dem Schlu, da in CelA die Bindung von β-1,4-Glucanen entweder auf der nicht reduzierenden Seite der -2 Substratunterbindestelle sterisch verhindert wird, oder alternativ zwei starke Substratunterbindestellen -1 und -2 vorhanden sind. Es wurde kein Signalpeptid zur Markierung des Proteins fr die Translokation ber die Membran gefunden. Zusammen mit der hohen Aktivitt gegenber Oligosacchariden, dem fast neutralen pH Optimum und der Inaktivierung bei niedrigem pH, deutet dies auf eine Rolle des Proteins als cytoplasmatisches Enzym zum Abbau importierter Oligosaccharide hin.

Ein zweites Enzym mit Xylanase- und Cellulaseaktivitt wurde zunchst aus *A. acidocaldarius* Kulturen gereinigt. CelB besa eine molekulare Masse von 100 kDa und konnte nur mit Hilfe von Detergenz von den Zellen abgelst werden. Klonierung und Sequenzanalyse des entsprechenden Gens ergaben einen offenen Leserahmen, welches fr ein Prprotein kodierte, das ein typisches Sec-abhngiges Signalpeptid besa. Gereinigtes, rekombinantes CelB und eine verkrzte Variante, der die letzten 203 Aminosuren fehlten, zeigten Enzymaktivitten, die dem Wildtypprotein hnlich waren. Ein niedriges pH-Optimum von 4 wurde bestimmt. Auch die Stabilitt war bei niedrigem pH hoch, wobei das Enzym nach Inkubation ber nacht bei pH 1.5 bis 7 eine Restaktivitt von 80 % aufwies. Das Temperaturoptimum betrug 80 C. Bei dieser Temperatur war das Enzym auch stabil und zeigte nach 1 h 60 % Restaktivitt. CelB hatte die Spezifitt eines Endoenzyms, jedoch wurden nach lngeren Inkubationszeiten Cellobiose und Xylobiose aus Cellulose bzw. Xylan freigesetzt. Das Enzym gehrte zur Familie 51 der Glykosylhydrolasen, aber es war erst der zweite Eintrag dieser Familie mit typischer Endoglucanaseaktivitt. CelB hatte hchste Sequenzhnlichkeit mit der zweiten Endoglucanase, EGF aus *Fibrobacter succinogenes*, wobei diese beiden Proteine eine markante Gruppe im phylogenetischen Baum dieser Familie bildeten.

Die Analyse der Aminosurezusammensetzung der katalytischen Domnen ergab, da CelB in bereinstimmung mit der Anpassung an einen niedrigen pH-Wert, weniger geladene Aminosuren als die neutrophilen Enzyme der gleichen Familie besitzt.

Wildtyp-CelB und unverkrztes, rekombinantes CelB waren nur in Anwesenheit von Detergenz lslich. Dagegen war das verkrzte CelB Protein vollstndig in Wasser lslich. Daher wird eine Rolle der C-terminalen Region bei der Zellassoziation nahegelegt. Dieser hydrophobe Bereich zeigte lokale bereinstimmungen der Aminosuresequenz mit einer hydrophoben Region einer Amylase aus dem gleichen Organismus.

3 Introduction

3.1 Aim of this work

Plants constitute by far the largest proportion of biomass on earth. Most plant carbon is present in the form of cellulose and hemicellulose. These compounds flow through the carbon cycle by degradation provided for by microorganisms. Thus, polysaccharide hydrolysis by bacteria is one of the most important enzymatic processes on earth (Schwarz (2001)).

Cellulolytic and hemicellulolytic activities are widespread in thermophilic microorganisms. Their occurrence is testimony to the presence of these substrates in thermophilic environments, either as plant litter in natural hot springs or in environments such as compost. Remarkably however, with a few exceptions, degradation of cellulose and hemicellulose among thermophiles is mostly due to anaerobic species and is absent in archaea (Bergquist et al. (1999)).

One of the first aerobic thermophiles to be described was *Alicyclobacillus acidocaldarius*, isolated in 1971 from a habitat that was not only hot but also very acidic. In the original description it was already noted that the organism was able to grow on starch (Darland and Brock (1971)). In my diploma thesis I could show that *A. acidocaldarius* was also able to grow on oat spelt xylan, a major component of hemicellulose, and began a tentative analysis of the xylanolytic activity (Eckert (1998)).

The goal of this work was to characterize and clone enzymes responsible for xylan degradation in the thermoacidophile *A. acidocaldarius*.

3.2 *Alicyclobacillus acidocaldarius*

Alicyclobacillus acidocaldarius was isolated by Darland and Brock (1971) from acidic hot springs (pH 2 to 3) in Yellowstone National Park (fig. 3.1). Enrichment was performed at a pH of 4.7 and a temperature of 55 C. In accordance with its surroundings optimal conditions of growth of the bacterium were found to be pH 3 and 65 C. *Alicyclobacillus acidocaldarius* is able to grow on a variety of sugars and polysaccharides as sole energy and carbon source, but not on organic acids. The isolate was described as a

Gram positive, endospore-forming, obligate aerobic rod and was therefore classified as belonging to the genus *Bacillus*. It was however noted that the GC content of 62 % was unusually high. Later, 16S rRNA studies demonstrated that three species belonging to aerobic, spore-forming bacteria represented a distinct phylogenetic group. This led to the creation of a new genus *Alicyclobacillus* (Wisotzkey et al. (1992)) with the three members *A. acidocaldarius*, *A. acidoterrestris* and *A. cycloheptanicus*. All three grow at high temperatures and low pH, *A. acidocaldarius* being adapted to the extremest conditions.

The genus name *Alicyclobacillus* refers to a further distinct trait, the presence of *ali*phatic, *cycl*ic fatty acids in the cytoplasmic membrane. These ω-alicyclic acids contain terminal cyclohexyl or cycloheptyl rings (fig. 3.2). Experiments with *A. acidocaldarius* mutants unable to synthesize cyclohexyl fatty acids demonstrated the importance of these lipids for growth at temperatures above 50 C and below pH 4. Studies on artificial membranes showed that addition of cyclic fatty acids lowered the transition temperature of membranes, but also had a condensing effect, leading to a higher impermeability for low molecular substances up to 20 C above the transition temperature. Thus, these lipids adapt membranes to extreme pH and temperatures (Deinhard and Poralla (1996)). They have been found in a variety of thermoacidophilic bacilli (Hippchen et al. (1981)). Interestingly, an isolate very similar to *A. acidocaldarius* was found in soil samples from neutral environments. Hippchen et al. (1981) suggest that acidic microenvironments exist in soil, making *A. acidocaldarius* more widespread than previously thought.

Electron micrographs of *Alicyclobacillus acidocaldarius* showed that the surface structure of the bacterium is composed of protein subunits arranged in a crystalline array, termed an S-layer (fig. 3.3). An S-layer is present in a large number of species that comprises all the major groups of bacteria (Engelhardt and Peters (1998)). It can play a protective role, since S-layer proteins are extremely resistant to harsh treatment. It may also participate in the adherence of certain bacteria to surfaces (Neidhardt et al. (1990)). S-layer proteins contain pores with varying diameters (2–8 nm) which can be more or less specific (Sleytr and Sra (1997)).

S-layer proteins from Gram-positive bacteria are bound to secondary cell wall polymers that are covalently linked to the peptidoglycan backbone. Conserved domains, called S-layer homology (SLH) domains are responsible for this interaction. They are found not only in S-layer proteins, but also in extracellular enzymes and proteins mostly involved in polysaccharide degradation and outer membrane proteins, where they also serve as an anchor to the surface of the bacterium. Therefore, the mechanism of anchoring SLH motifs interacting with secondary cell wall polymers seems widespread among prokaryotes and has been conserved during evolution (Sara (2001),

Figure 3.1: A hot spring in Yellowstone National Park (USA). The steaming spring with its extreme environment gives way to mesophilic vegetation as the distance increases. E. Schneider, with permission.

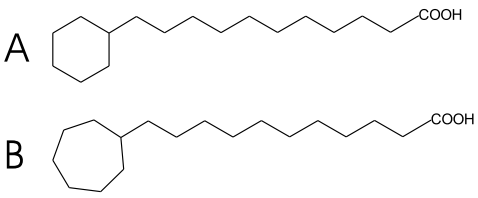


Figure 3.2: Structure of aliphatic, cyclic fatty acids found in the cytoplasmic membrane of *Alicyclobacillus* species. (A) ω-cyclohexylundecanoic acid **(B)** ω-cycloheptylundecanoic acid. Deinhard and Poralla (1996).

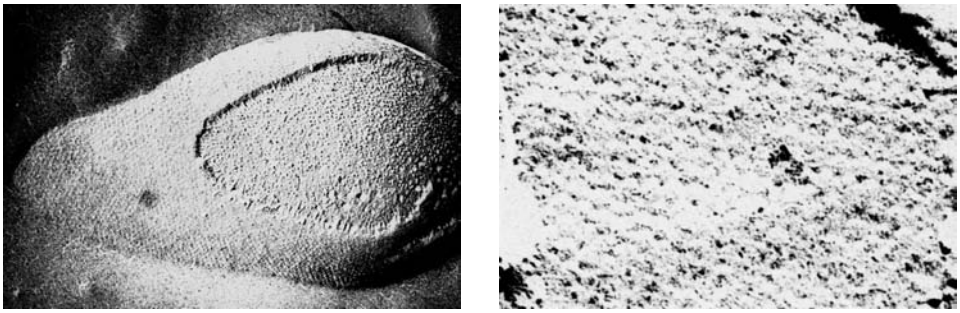


Figure 3.3: Electron micrographs of *A. acidocaldarius*. *Left:* Cell after freeze etching. Note the S-layer. *Right:* Detail of the S-Layer with hexagonal lattice. The centre to centre distance is 13.8 nm. Messner (1994), with permission.

Engelhardt and Peters (1998)).

A gene cluster containing genes involved in starch utilization has been found in *A. acidocaldarius* (Hlsmann et al. (2000a), Hlsmann (2000b)). A putative operon contains genes coding for three glycoside hydrolases involved in the breakdown of starch, and also an active transport system for maltose and maltodextrins.

Of the three starch degrading enzymes (an amylase, a cyclomaltodextrinase and a glucosidase), two enzymes have been described in greater detail. The amylase has been purified, and the gene *amyA* cloned and sequenced. Interestingly, it remained cell-bound during logarithmic growth, but was released in the stationary growth phase. The cell-associated form was extractable from cells with the help of detergent (Schwermann et al. (1994)). It displayed a pH and temperature optimum of 3 and 75 C, respectively. The corresponding gene (*amyA*) coded for a protein with a calculated molecular mass of 140 kDa (Koivula et al. (1993)). The wild-type protein was found to be highly glycosylated, displaying a molecular mass of 160 kDa as determined by SDS polyacrylamide gel electrophoresis. How cell surface attachment was mediated remained unclear.

More recently the cyclomaltodextrinase (CdaA) was subject to research. It degraded cyclomaltodextrins and pullulan. The enzyme displayed a high temperature optimum but a near neutral pH optimum which was in accordance with its localization in the cytoplasm (Matzke (1999), Matzke et al. (2000)).

Little is known about the genetics of *A. acidocaldarius*. A plasmid has been reported, but the function is unknown (Decker (1987)). To date genetic transformation of the organism has not been achieved.

3.3 Adapting proteins to the extreme

Strikingly, acidophiles do not adapt the entire cellular machinery to acidic conditions. Instead they maintain a near neutral intracellular pH. The proton motive force of these bacteria relies on a large ΔpH and a low or even reversed membrane potential. This is probably achieved by alkali cation uptake systems which move positive charges into the cell, lowering the membrane potential (Bakker (1990)). Since it is unnecessary to adapt intracellular proteins to low pH, only the components of the cell which are exposed to the acidic environment need to be acidophilic. These include extracellular enzymes and exposed regions of membrane proteins.

Even though acidophilic proteins have been cloned and characterized (Inagaki et al. (1998), Ruiz et al. (1997), Kimura et al. (2000), Hlsmann et al. (2000a)), the mechanism of protein adaptation to low pH remains somewhat unclear. Especially striking are cases of enzymes which show high similarities in their primary sequences, but differ in pH optimum and stability. For example, a thermostable isomerase with an acidic pH optimum from *Thermoanaerobacterium* (Liu et al. (1996)) was compared with two other thermostable isomerases. Except for the low pH optimum it had similar biochemical characteristics to those of two other isomerases. The primary sequence differed only in 7 or 10 amino acids over a length of 439 residues. This was enough to shift the pH optimum by 0.5 units. A similar situation has been described for an acidophilic xylanase (Ohta et al. (2001)). All acidophilic enzymes described, showed overall sequence similarities to mesophilic enzymes. Together this points towards slight changes in proteins for adaption to low pH instead of a radical new design. A difficulty hampering research in this area is that two enzymes will differ in more than one respect. Putting down differences in amino acid composition to a certain property then becomes speculative.

It must also be remembered that when determining an optimum parameter for a protein two characteristics play an important role: activity and stability. A low stability, i.e. quick denaturing of a protein, will lead to a low indication of activity under the same conditions especially when measurements are carried out over a long period of time. Whereas activity may be influenced by subtle changes to the active site, stability is a trait of the overall protein structure.

Despite the difficulties mentioned, hypotheses have been put forward, explaining acidostability. One theory, proposed by Schwermann et al. (1994) suggests a lower charge density in acidophilic proteins as the reason for stability. In neutrophilic proteins exposure to low pH leads to protonation of all basic and acidic residues. Basic residues remain positively charged, whereas the carboxyl groups of the acidic amino acids become uncharged. This results in an excess positive charge, leading to denaturing of the pro-

tein. In acidophilic proteins this is circumvented by reducing positive and negative charges. Indeed, extracellular proteins from acidophiles including *A.acidocaldarius* were found to have a low charge density, especially on the surface of the protein.

Acidophilic proteins, must not only remain stable under acidic conditions, they must also display activity at low pH. For enzymes this implies a functional active site. In particular, catalytic amino acids must be in the right protonation state. This has been extensively studied in certain acidophilic xylanases. In these enzymes two carboxylic acid residues are responsible for catalysis, one serving as an acid-base catalyst. In acidic xylanases a neighbouring aspartate residue is within hydrogen bonding distance. It has been proposed that at low pH this residue becomes protonated, making the proton of the acid-base catalyst available for catalysis. Xylanases with an asparagine residue at this position invariably have a higher pH optimum, since the asparagine side-chain can always serve not only as a hydrogen bond acceptor, but also as a hydrogen bond donor at higher pH (Krengel and Dijkstra (1996)). Fig. 3.4 shows the interactions of the acid-base catalyst as found in the protein crystal.

Single-celled organisms like bacteria cannot shield themselves against harsh temperatures. Temporary adaptation can take place by producing heat-shock proteins (acquired thermotolerance), but to cope with elevated temperatures for prolonged amounts of time all components of the cell must be adapted to high temperatures.

In an 'average' protein the sum of stabilizing interactions is large, about 1 MJ·mol^{-1}. Destabilizing forces are also large, but $\Delta G_{N\rightarrow U}$, the free energy of stabilization, i.e. the difference between the two, is only of the order of 40 kJ mol^{-1}. When comparing $\Delta G_{N\rightarrow U}$ for mesophilic and thermophilic proteins, small increases of $\sim$ 60 kJ mol^{-1} are found. This corresponds to only a few extra stabilizing interactions (Wassenberg et al. (2000), Daniel et al. (1996)). Thus, comparing the 3 dimensional structures has led to the conclusion that, as in acidophilic proteins, the structures of homologous proteins from thermophilic sources are generally very similar to their mesophilic counterparts. The increased numbers of interactions in thermophilic proteins are constituted of the generally known mechanisms of protein stabilization: increased van der Waals interactions, H-bonds, salt bridges, dipole-dipole interactions, disulfide bridges, and hydrophobic interactions (Kumar et al. (2000), Aguilar et al. (1997), Rice et al. (1996), Natesh et al. (1999)). However, elucidating which differences are directly responsible for thermostablility has proven difficult, because proteins seldom differ only in this respect.

The drawback of using a structural / mutational approach to the problem of thermostability lies in the labor intensity which restricts the number of proteins being studied. This results in a potentially biased view of ther-

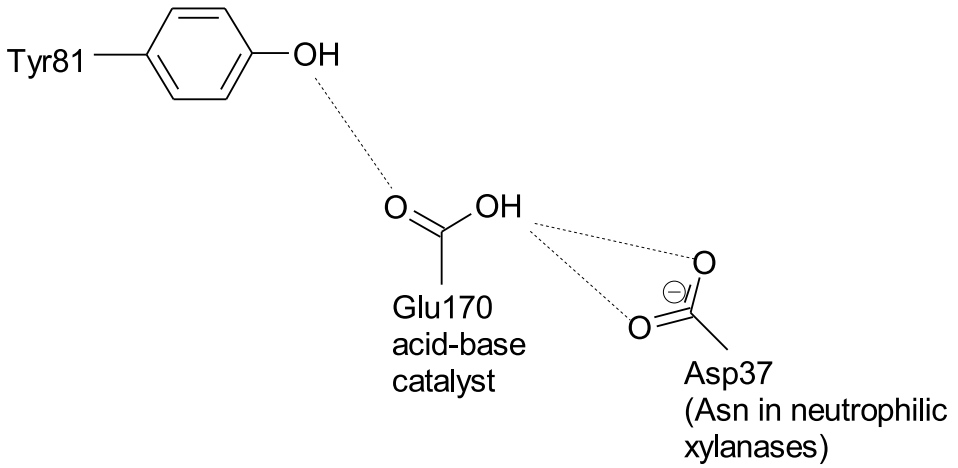


Figure 3.4: Sketch of the interactions of the acid-base catalyst of *A. niger* xylanase I at pH 8. Dotted lines denote hydrogen bonds. At low pH protonation of the aspartate residue may make the proton of the catalytic residue available for catalysis. Taken from Krengel and Dijkstra (1996).

mal stabilizing mechanisms. A second approach uses sequence comparisons of families of homologous high- and low-temperature proteins. Here statistical analyses extract recurring amino acid replacement trends (Haney et al. (1999), La et al. (2003)). Using such analyses it has been found that the observed replacements decrease the content of uncharged polar residues, increase the content of charged residues and increase residue hydrophobicity and residue volume for the thermophilic proteins relative to their mesophilic counterparts (Haney et al. (1999)). This leads to more salt bridges and a tighter packing of the hydrophobic core of the protein.

An interesting feature described for some proteins have been so-called thermostabilizing domains. These protein domains have been hypothesized as functioning to increase the stability of a protein with respect to adverse environmental conditions like high temperatures, low pH or proteases (Hayashi et al. (1997), Lee et al. (1993), Clarke et al. (1996)). However, these assignments should be viewed with caution, because functional analysis of such domains has proved misleading. Removal of a domain resulting in thermolability of the truncated protein has been taken as evidence for a role in thermostabilization. More convincingly, transfer of such a domain to a mesophilic protein has also resulted in higher thermostability. Later studies ascribed a different function, namely xylan binding to such domains, stabilization then being seen as a side-effect (Meissner et al. (2000)).

3.4 Plant cell walls

The protoplast of plant cells is surrounded by a cell wall. It is the exoskeleton of the cell and gives it form and rigidity. Structurally speaking it contains a fibrillar scaffolding, the cellulose fibrils, and an amorphous matrix made up of hemicellulose, pectins, proteins and phenolics (fig. 3.5). The composition of individual cell walls can vary widely depending on cell type (Brett and Waldron (1996)).

The secondary cell wall is deposited after the cell wall has reached its final surface area and plays a major role in the structural rigidity of woody plants. It contains a lot more cellulose than the primary cell wall. Secondary cell walls of wood typically consist of two thirds cellulose and hemicellulose (mainly xylan) and one third lignin, a phenolic compound (Sitte et al. (1991)). The structure of the major polysaccharides cellulose and xylan shall be discussed here in greater detail.

Cellulose is the most abundant biopolymer on earth, with $\sim 10^{12}$ t being produced in the biosphere per year. It is a polysaccharide made up of D-glucose units linked by 1,4-β-D-glucosidic bonds with a degree of polymerization (DP) of 2000–15,000. Within a single chain neighbouring glucose units are rotated by 180 with respect to each other. This results in an elongated structure, shown in fig. 3.6, with hydrogen bonds forming between the ring oxygen of one glucose residue and the C3 hydroxyl hydrogen of an adjacent residue.

In cell walls single chains associate via H-bonds to form microfibrills. These fibrills are of a mainly crystalline nature, but amorphous regions are also present. The crystalline nature of cellulose fibrils makes them highly resistant to enzymatic degradation (Sitte et al. (1991)).

Xylan, the major component of hemicellulose, is a complex polysaccharide made up of a backbone of β-1,4-linked D-xylose units which carries various substituents. The main chain can be acetylated or substituted with glucuronic acid in addition to which it can itself be branched, carrying arabinofuranose or 4-*O*-methyl-D-glucuronic acid side chains (Thomson (1993), fig. 3.7). Because xylose, like glucose, forms a pyranose ring the two sugars are structurally similar (fig. 3.8). Xylan adopts an elongated structure as does cellulose. However, xylan exists as a helical structure at least under crystallized conditions (Kulkarni et al. (1999)). The interactions between xylan and cellulose in the plant cell wall are thought to be non-covalent.

Xylans from different sources exhibit considerable variation in composition. Hardwood xylan is acetyl-4-*O*-methylglucuronoxylan with a DP of ~ 200. Softwoods contain arabino-4-*O*-methylgucuronoxylan which is not acetylated and has a DP of ~ 100. Finally, the xylan of grasses and cereals is also arabino-4-*O*-methylgucuronoxylan with a DP of ~ 70. It has less glucuronic acid, but does contain a large content of arabinofuranosyl side

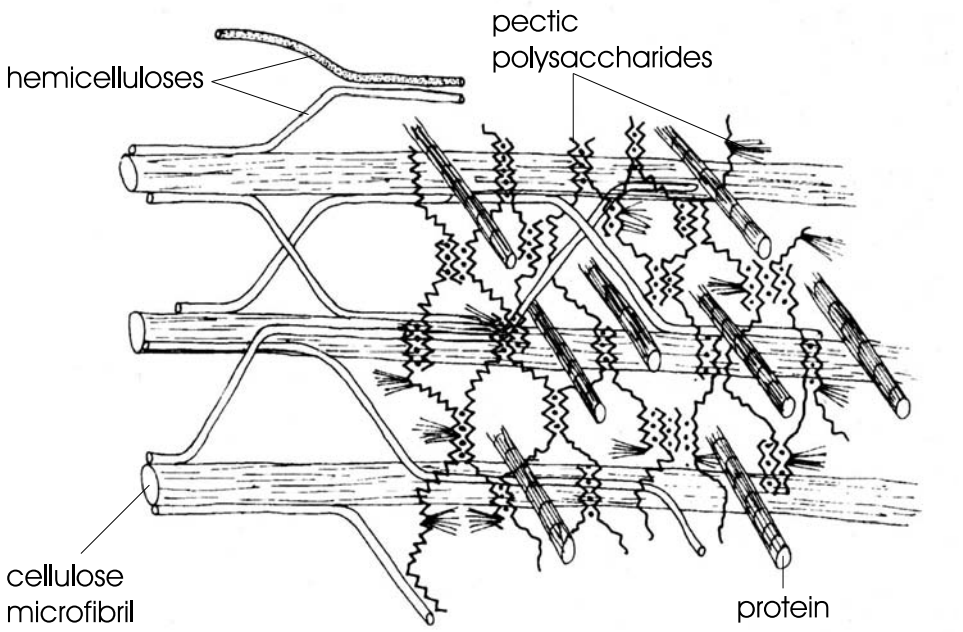


Figure 3.5: Schematic representation of the plant cell wall and the interactions of its polymers. Taken from Brett and Waldron (1996)).

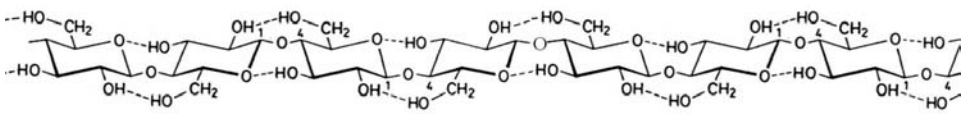


Figure 3.6: Structure of a β-1,4-glucan as found in cellulose. Note the H-bonds between adjacent glucose units. Adapted from Sitte et al. (1991).

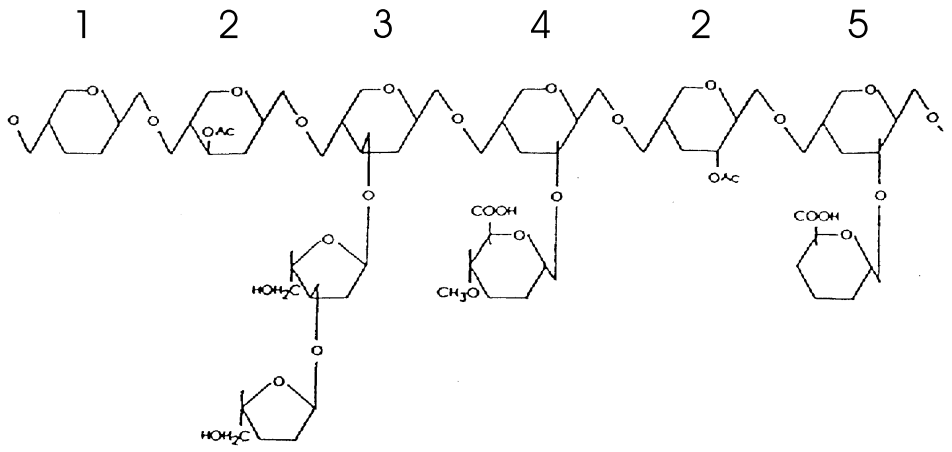


Figure 3.7: Part of a hypothetical xylan molecule. *1*, xylose; *2*, acetyl xylose; *3*, α-L-arabinofuranose side chain; *4*, 4-*O*-methyl-D-glucuronic acid side chain; *5*, D-glucuronic acid. Taken from Thomson (1993).

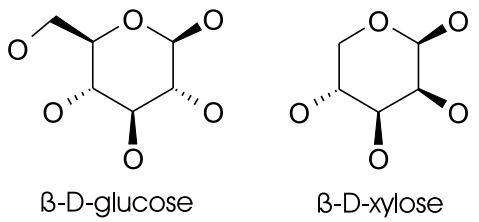


Figure 3.8: Structures of β-D-glucopyranose and β-D-xylopyranose. Although glucose is a C6 sugar and xylose a C5 sugar both exist in the pyranose form.

chains and is partly acetylated (2–5 % of dry matter, representing acetylation of half of the xylose units). Moreover, ∼ 5 % of the arabinosyl side chains are themselves substituted with feruloyl or coumaroyl groups (Bacon et al. (1975), Coughlan and Hazlewood (1993), Kulkarni et al. (1999), Thomson (1993)). Tab. 3.1 gives an overview of the different xylans.

From a structural point of view the substituents can influence the interaction between xylan and cellulose in the plant cell wall, the side-chains preventing hydrogen bonding between the two polymers (Brett and Waldron (1996)). As far as microbial attack is concerned, xylan degradation requires multiple enzymes due to the heterogeneity of the substrate.

Source	DP	Gluc. acid	Acetylation	Other
glucuronoxylan				
hardwood	150–200	10 %	70 %	
arabinoglucuronoxylan				
softwood	70–130	20 %	none	13 % arabinose
cereals, grasses	70	low	50 %	large content of arabinose, feruloyl or coumaroyl groups

Table 3.1: Composition of xylan from various sources. Given is the percentage of substituted xylose residues for the different substituents. *Gluc. acid* is 4-*O*-methyl-α-D-glucuronic acid. Data from Bacon et al. (1975), Coughlan and Hazlewood (1993), Kulkarni et al. (1999), Thomson (1993).

3.5 Glycoside hydrolases

The recycling of photosynthetically fixed carbon is a key microbial process. Because carbohydrates can vary in the nature of the monomers, the linkages, chain length and substituents, microorganisms are faced with a huge variety of sugars which is mirrored by the many different *O*-glycoside hydrolases (EC 3.2.1.x).

In contrast to the many different specificities, hydrolysis of the scissle bond takes place by only two catalytic mechanisms, inverting and retaining, which proceed via inversion or retention of the configuration at the anomeric carbon C1 of the sugar unit.

For retaining glycoside hydrolases (GH) catalysis proceeds via a double displacement mechanism shown in fig. 3.9 and involves a covalent glycosyl-enzme intermediate (Koshland (1953)). One active site residue acts as a nucleophile, attacking the anomeric carbon of the substrate to displace the

leaving group and form the covalent intermediate. The other residue plays the dual role of acid / base catalyst. It assists in the formation of the covalent intermediate by proton donation to the leaving group of the departing aglycone. In the subsequent hydrolysis step it functions as a base catalyst, deprotonating the attacking water. Both steps of the mechanism go through high-energy oxocarbenium-like transition states, in which the reaction centre of the sugar is planar and positively charged (Uitdehaag et al. (1999)). The catalytic residues, typically aspartic or glutamic acids, are positioned ~ 5.5 Å apart in the active site (Lawson et al. (1997)).

Many retaining enzymes also catalyze glycosyl transfer between two substrate molecules (Sinnott (1991)). In this case the glycone moiety (the glycosyl group of the covalent intermediate) is transferred not to water, but to a hydroxyl group of a second sugar residue. Such transferases (EC 2.4.x.y) can play a role in the synthesis of complex polysaccharides, e.g. glycogen or starch (Harjunp et al. (1999), Meissner and Liebl (1998), Withers (2001)). Usually the transfer takes place from a specific donor to a relatively large number of structurally different acceptors (Park et al. (1998)). If transglycosylation occurs within the same molecule (intramolecular transglycosylation) cyclic structures can arise (Przylas et al. (2000)). How enzymes control hydrolysis versus transglycosylation still remains somewhat unclear, however higher substrate concentrations favour transglycosylation (Saloheimo et al. (2002)).

Catalysis with inversion of configuration shown in fig. 3.10 proceeds via a single nucleophilic displacement. In this reaction hydrolysis of a β-glycosidic bond creates a product with the α-configuration and vice versa. Again, two catalytic carboxylates are positioned on opposite sides of the sugar ring. One functions as general acid, catalyzing leaving group departure. The other acts as a general base which activates a water molecule by extracting a proton, which leads to attack of the anomeric carbon of the scissle bond. This attack from the backside leads to inversion of the absolute configuration (Sinnott (1991)). Catalytic residues are typically further apart ($\sim$ 10–11 Å) than in retaining enzymes (Lawson et al. (1997)).

While the catalytic action of glycoside hydrolases is well understood in general terms, details of the interactions between the glycosidic chain and the protein are less well characterized. However, this governs the specific aspects of enzyme function, i.e. endo or exo mode of action, minimum size of oligomer which is still cleaved, reactivity towards branched substrates, etc. (Schmidt et al. (1999)).

Three different types of active site topologies have been found in glycosyl hydrolases (Gruber et al. (1998)): (*i*) Pockets are optimal for small substrates and enzymes adapted to substrates with a large number of available chain ends. (*ii*) Tunnels require the polysaccharide chain to be threaded through. This works best for unbranched chains. All known exocellulases,

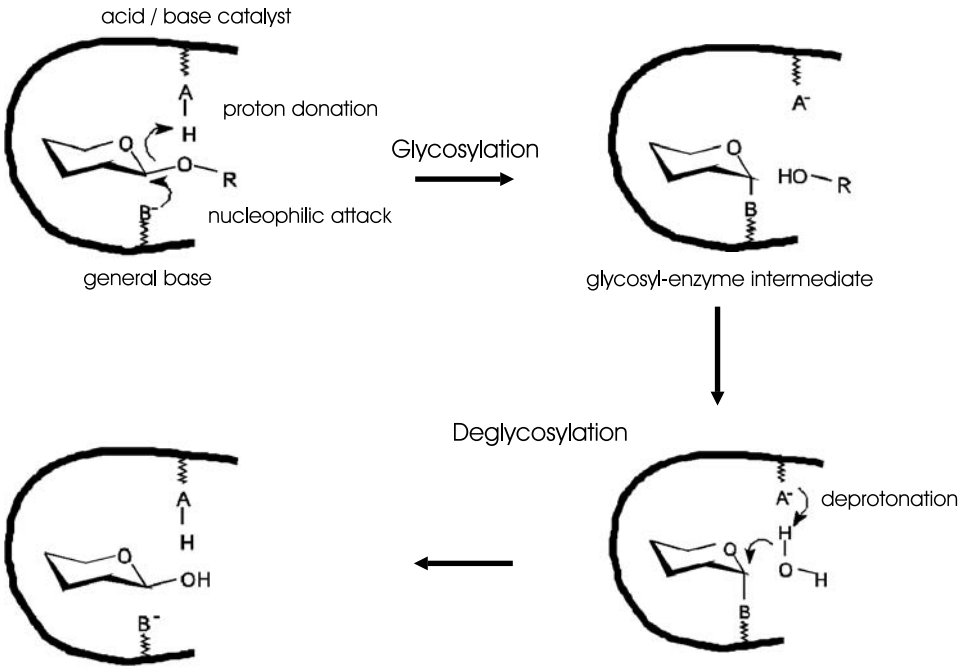


Figure 3.9: Catalytic mechanism of retaining glycoside hydrolases. In the first step a glycoside-enzyme intermediate is formed, followed by deglycosylation. In glycoside transferases the attacking water molecule is replaced by a sugar hydroxyl. Taken from the CAZy web site (http://afmb.cnrs-mrs.fr/CAZY/GH.html).

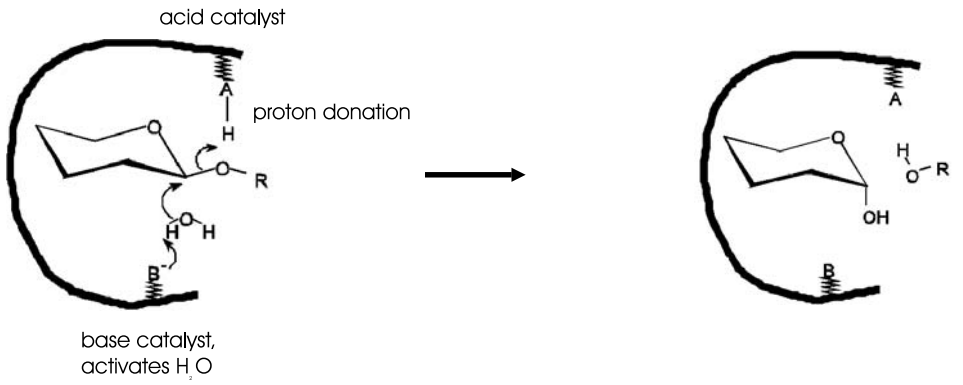


Figure 3.10: Catalytic mechanism of inverting glycoside hydrolases. Catalysis with inversion of configuration proceeds via a single nucleophilic displacement. Taken from the CAZy web site (http://afmb.cnrs-mrs.fr/CAZY/GH.html).

which hydrolyze cellobiose units from the end of a cellulose chain have such a topology (Rouvinen et al. (1990), Sakon et al. (1997)). (*iii*) Clefts or grooves are open catalytic sites and allow the binding of several consecutive sugar units in linear or branched polymeric substrates. They are commonly observed in endo-acting enzymes (Schmidt et al. (1998)).

Sugar subunits are held in position by hydrogen bonds with their hydroxyl groups as well as by hydrophobic stacking interactions between the hydrophobic faces of the sugar rings and the aromatic side chains of tryptophans or tyrosines (Schmidt et al. (1999)). The binding sites of glycosyl hydrolases comprise a series of subsites, each one capable of binding a monosaccharide unit (Charnock et al. (1998), Hiromi et al. (1983)). Enzymes differ widely in number and affinity of the individual subsites. When describing enzyme topology, subsites are named from $-n$ to $+n$, with $-n$ at the non-reducing end and $+n$ at the reducing end. Cleavage occurs between the -1 and $+1$ subsites (Davies et al. (1997)), so the subsites that bind the glycone and aglycone region are prefixed by $-$ and $+$, respectively (fig. 3.11).

The IUB-MB (International Union of Biochemistry and Molecular Biology) nomenclature with EC (Enzyme Commission) numbers classifies enzymes according to their specificity and the type of reaction catalyzed. This is useful for providing a unique classification, but becomes problematic with enzymes such as glycoside hydrolases that exhibit a broad substrate specificity. A second system based on sequence similarities and hydrophobic cluster analysis has been proposed (Bourne and Henrissat (2001), Henrissat (1991), Henrissat et al. (1995)). To date 91 families have been described.

The new system intended to reflect evolutionary relationships and structural similarities, which EC numbers were never meant to do (Henrissat (1991), Henrissat et al. (1998)). It must be pointed out that only small differences in the primary sequences of enzymes can have effects on the active site topology producing profound changes in substrate specificity and even the cleavage mechanism (Juy et al. (1992), Natesh et al. (1999), Przylas et al. (2000), Sakon et al. (1996)). Thus, such divergent evolutionary events or convergent evolution forcing enzymes with different folds to catalyze the same reaction are better mirrored by a sequence based approach, which keeps related enzymes in the same group (Henrissat and Bairoch (1993)). Such a group is then polyspecific, containing enzymes with different specificities.

Several families have been grouped into clans. A clan is a group of families that are thought to have a common ancestry and are recognized by significant similarities in tertiary structure (Henrissat and Bairoch (1996)). Because all glycoside hydrolases found up to date act by a general acid catalysis mechanism and the position of the two catalytic residues is conserved in related enzymes, sequence analyses may be used to predict the active site residues (Henrissat (1991), Henrissat et al. (1998)).

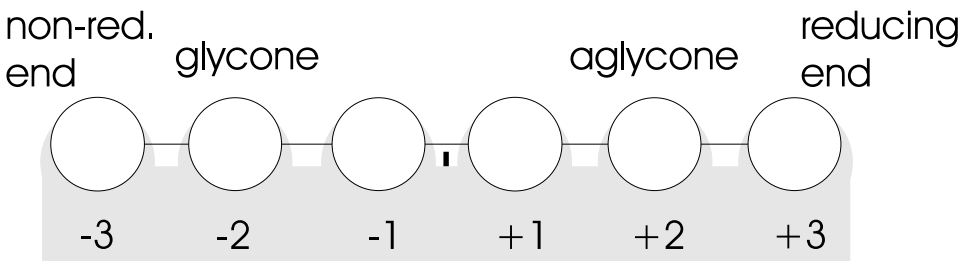


Figure 3.11: Diagram indicating the nomenclature for sugar-binding subsites in glycosyl hydrolases. The enzyme is in *gray*, the position of the cleavage site is indicated by the *black bar*. Subsites are *numbered* according to Davies et al. (1997). Sugar units are represented by *circles*. The non-reducing and the reducing end of the sugar are given as well as the glycone and aglycone moieties.

An update of the classification of glycoside hydrolases is provided by the Glycoside Hydrolase Family Server at the 'Carbohydrate-Active enZYmes' website (CAZy, sec. 4.6, Coutinho and Henrissat (1999)).

3.5.1 Domain organization of glycoside hydrolases

Many microbial sugar degrading enzymes are multidomain proteins. Apart from the catalytic domain they contain further elements with a variety of functions. These can be elements involved in attaching the enzyme to the substrate, i.e. cellulose or xylan binding domains (Gilkes et al. (1991)), domains involved in cell-surface attachment or even further catalytic domains with a second substrate specificity that act in synergy with the other (Fernandes et al. (1999), Laurie et al. (1997)).

Domains are often connected by characteristic linker sequences ($\sim$ 20–30 aa) which may offer the necessary flexibility between domains, but may also play a role analogous to that of introns in eukaryotes by allowing domain shuffling to occur more easily (Kulkarni et al. (1999)).

An example of domains involved in cell association are the aforementioned SLH domains (sec. 3.2). Another example is provided by the dockerin domains of *Clostridium thermocellum* that are responsible for the binding of various proteins to an extracellular, cell-bound cellulose-degrading complex, the cellulosome. In this 'organelle' the dockerin domains interact with the cohesin domains of a non-catalytic scaffolding component CipA (Beguin and Alzari (1998)).

Domains which attach the enzyme to the surface of the polysaccharide are also known as carbohydrate binding modules (CBMs). Like catalytic domains they are classified in families on the basis of amino acid sequence similarities (Tomme et al. (1995)). Interaction with the substrates is also mediated by hydrophobic stacking. Interestingly they have varying affinities

for their substrates. Some can be removed from insoluble substrates by treatment with water (Cann et al. (1999)) or sugar solutions (Sun et al. (1998), Takada et al. (1996)), others require heating in 5 % SDS (Lopez-Fernandez et al. (1998), Millward-Sadler et al. (1995)). The exact function of CBMs is still being elucidated. Some merely bind to the surface of the substrate and increase the concentration of the catalytic domain directly on the substrate (Fernandes et al. (1999), Rouvinen et al. (1990)) whilst others actually seem to disrupt the structure of the insoluble material and may even make individual chains accessible to the catalytic domain (Gill et al. (1999)).

3.5.2 Degradation of cellulose and xylan

The structural homogeneity of cellulose is in stark contrast with the complex catalytic machinery needed to degrade it. This is because the crystalline nature of the substrate results in a structure highly resiliant to enzymatic attack (Beguin (1994)). CBMs are needed for the attachment of the catalytic components on the surface of the substrate. These may make the polysaccharide more susceptable to attack by catalytic domains. Endoglucanases (EC 3.2.1.4) can then degrade soluble and amorphous regions and produce new ends by cutting the strands. Exoglucanases (EC 3.2.1.74) or cellobiohydrolases (EC 3.2.1.91) attack the ends progressively. Finally, β-glucosidases (EC 3.2.1.21) degrade the resulting oligosaccharide products (Schwarz (2001)). Endoglucanases have been found in GH families 5–10, 12, 44, 45, 48, 51, 61 and 74.

Anaerobic bacteria possess a multi-enzyme system called the cellulosome to meet the challenge of cellulose degradation, whereas to date aerobic bacteria have been found to express individual extracellular enzymes (Schwarz (2001)).

Xylan does not adopt a crystalline packing, but it is insoluble and enzymatic degradation is further complicated by the many substituents. Main chain cleaving enzymes, i.e. endo-1,4-β-xylanases (EC 3.2.1.8) and β-xylosidases (EC 3.2.1.37), are often prevented from hydrolyzing the backbone by the presence of substituents. These must be removed before extensive degradation of the polymer can take place. Side chains are cleaved off by acetyl esterases (EC 3.1.1.6), α-L-arabinofuranosidases (EC 3.2.1.55) and α-glucuronidases (EC 3.2.1). Xylanases have been found in GH families 10, 11 and 43.

It is interesting to note that like lignin some bacteria degrade xylan to gain access to cellulose, but do not use it as carbon source. For example *Clostridium thermocellum* produces at least six xylanases, but cannot grow on xylose (Fernandes et al. (1999)).

4 Materials and methods

Composition of reagents in % is always given as v/v in liquids or w/v with solids in liquids. Ratios (e.g. 1 : 2) are always given as $v : v$. Nucleotide sequences are always given from 5' to 3', amino acid sequences are given from the N to the C-terminus.

4.1 Bacterial strains, phages, plasmids, and growth conditions

Tables 4.1 and 4.2 list the strains, plasmids and phage vectors used in this work.

Alicyclobacillus acidocaldarius ATCC 27009 minimal medium contained the following basal salts (stored as 40× stock solutions):

9.8	mM	$(NH_4)_2SO_4$
2.7	mM	KH_2PO_4
1	mM	$MgSO_4$
0.47	mM	$CaCl_2$

adjusted to pH 3.5 with HCl.

After autoclaving, 10 M $FeCl_3$ (1 mM stock solution, autoclaved seperately) and the following trace element salts prepared as a 100× concentrated stock solution and brought to pH 2 with H_2SO_4 were added:

11.7	M	$Na_2B_4O_7$
9.1	M	$MnCl_2$
0.57	M	$ZnSO_4$
0.2	M	$CuCl_2$
0.16	M	Na_2MoO_4
0.12	M	$VOSO_4$
30	nM	$CoSO_4$

Carbon sources (at 0.2 % each) were oat spelt xylan, birchwood xylan (Roth, Germany), starch (Sigma, Germany), sugar beet arabinan (Megazyme, Ireland), CMC (Serva Feinbiochemica, Germany) or glycerol. Oat

Strain	**Relevant genotype / description**	**Reference / source**
Escherichia coli		
JM109	e14$^-$(*mcrA*) *recA1 endA1 gyrA96 thi-1 hsdR17* (*rk*$^-$, *mk*$^+$) *supE44 relA1* Δ(*lac-proAB*) [F'*trad36 proAB lacI*^{q}*ZΔM15*]	Yanish-Perron et al. (1985)
TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*ΔM15 Δ*lacX74 deoR recA1 araD139* Δ(*araA-leu*)7697 *galU galK rpsL endA1 nupG*	Invitrogen
XL1-Blue MRF'	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F' *proAB lacI*^{q}*ZΔM15* Tn10(Tet^r)]	Stratagene
XLOLR	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 thi-1 recA1 gyrA96 relA1 lac* [F' *proAB lacI*^{q}*ZΔM15* Tn10(Tet^r)] Su$^-$ (nonsuppressing) λ^r (lambda resistant)	Stratagene
Alicyclobacillus acidocaldarius		
ATCC 27009	wild-type	Darland and Brock (1971)

Table 4.1: Strains used in this work with relevant genotypes and references.

Vector	Description (host vector)	Reference / source
pUC18	pBR322 derived cloning vector *lac lacZ* provides α-complementation	Yanish-Perron et al. (1985)
pBAD/HisA, B	arabinose dependent plasmid expression vectors, provide N-term. His_6-tag	Invitrogen
ZAP Express	λ derived cloning vector *lac lacZ* provides α-complementation	Stratagene
pBK-CMV	ZAP Express derived phagemid vector	Stratagene
ExAssist	f1 derived helper phage, allows *in vivo* excision of the pBK-CMV vector from the ZAP Express vector in the XLOLR strain	Stratagene
pKE1	plasmid isolated from the original *celA* clone from the *A. acidocaldarius* plasmid gene bank (pUC18)	sec. 5.1.1, fig. 5.3
pKE2	pKE1 subclone (pUC18)	sec. 5.1.2, fig. 5.3
pKE3	pKE1 subclone (pUC18)	sec. 5.1.2, fig. 5.3
pKE4	pKE1 subclone (pUC18)	sec. 5.1.2, fig. 5.3
pKE5	pKE1 subclone (pUC18)	sec. 5.1.2, fig. 5.3
pFZ1	*celA* clone (pBAD/HisA)	sec. 4.2.8, fig. 5.3, 5.4
pKE25	phagemid isolated from the original *celB* clone from the *A. acidocaldarius* ZAP Express gene bank (pBK-CMV)	sec. 4.2.10, fig. 5.17
pKE25a5	plasmid containing a *Kpn*I fragment of chromosomal DNA from *A. acidocaldarius* overlapping with the insert of pKE25 (pUC18)	sec. 4.2.10, fig. 5.17
pKE2201	$celB_{trunc}$ clone (pBAD/HisB)	sec. 4.2.11, fig. 5.17
pKE101	*celB* clone (pBAD/HisB)	sec. 4.2.11, fig. 5.17

Table 4.2: Plasmid and phage vectors used in this work with descriptions and references.

spelt xylan was added from a 2 % stock solution prepared by boiling for 10 s in basal salts. Maltose, arabinose (Roth, Germany), cellobiose, glucose or xylose (Merck, Germany) were added to a final concentration of 10 mM.

Complex medium consisted of minimal medium with glucose as carbon source supplemented with 0.1 % bacto tryptone and 0.1 % yeast extract. Cultures were incubated at 57 C with shaking at 260 rpm in a water bath shaker.

Escherichia coli strains were grown in Luria-Bertani (LB) medium (1 % tryptone, 0.5 % yeast extract, 1 % NaCl adjusted to pH 7 with NaOH) containing 100 g ml^{-1} ampicillin if appropriate. Cultivation took place at 37 C and aeration (180 rpm). LB plates, incubated at 37 C, contained an additional 1.5 % agar (Sambrook et al. (1989)).

Normal plates were 9 cm in diameter, large plates were 14 cm wide.

The λ Zap Express phage was plated, titered, amplified and stored according to the manufacturers instructions (Stratagene). For short term storage phages were kept at 4 C in SM buffer (Stratagene) with a drop of chloroform added to 1 ml.

4.2 Molecular methods

4.2.1 General DNA modification, electrophoresis and concentration measurement

DNA manipulations were done according to standard protocols (Sambrook et al. (1989)). Restriction enzymes were from NEB, shrimp alkaline phosphatase was from Amersham-Buchler and T4 DNA ligase was from Boehringer.

DNA segments were seperated by agarose gel electrophoresis according to (Sambrook et al. (1989)). To isolate segments from gels the Gibco Concert gel extraction system was employed.

Prior to ligation DNA segments were always purified via preparative gel electrophoresis and vectors were always dephosphorylated. Ligation took place over night at 16 C when ligating staggered ends and at 4 C for blunt ends.

Site-directed mutagenesis was carried out using Stratagene's QuikChange kit.

DNA concentration was assayed using a Pharmacia Ultrospec 2000 photometer according to the manufacturer.

4.2.2 Southern blotting

DNA was transferred from agarose gels to nylon Hybond N^+ membrane (Amersham Buchler) by Southern blotting using a vacuum transfer LKB

2016 VacuGene apparatus (Pharmacia) according to the manufacturers instructions. After transfer and UV cross-linking hybridization and detection took place over night as described in Eisel et al. (2000). Probes were either purchased as 5' digoxigenin (Dig) labelled oligonucleotides from MWG or labelled with the Dig-DNA labelling and detection kit (Boehringer) and the amount of labelled DNA quantified. Probe-target hybrids were detected with anti-Dig-alkaline phosphatase antibodies and the chromogenic substrate nitroblue tetrazolium / 5-bromo-4-chloro-3-indolyl phosphate.

4.2.3 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was prepared either 'by hand' (Morelle (1989)) or using the 'Mini / Midiprep Plasmid Kit' (Qiagen).

4.2.4 Isolation of chromosomal DNA from *A.acidocaldarius*

Chromosomal DNA was isolated according to a modified protocol of Harwood and Cutting (1990) and Hlsmann (2000b) in the following manner:

200 ml of an *A. acidocaldarius* culture was grown over night ($\sim$ 20 h) in complex medium to an OD_{650} of $\sim$ 1.5. After cooling briefly on ice cells were harvested by centrifugation for 10 min at $8,000 \times g$ and treated with 40 mg of lysozyme for 25 min at 37 C in 8 ml lysis buffer (50 mM EDTA, 0.1 M NaCl, pH 7.5). Following addition of *N*-lauroylsarcosine to a final concentration of 1.5 % incubation continued for 5 min. All subsequent steps took place on ice or at 4 C. DNA was extracted by shaking three times with the same volume of phenol and twice with the same volume of phenol / chloroform / isoamyl alcohol (24 : 24 : 1) between which the aqueous phase was recovered by centrifugation at $10,000 \times g$. After a final chloroform / isoamyl alcohol (24 : 1) extraction chromosomal DNA was precipitated by adjusting the sample to 0.3 M Na acetate (pH 5.5), adding 2.5 volumes of ethanol (-20 C), storing at -20 C for 30 min and centrifuging for 10 min at $10,000 \times g$. The DNA pellet was then washed with 70 % ethanol and dissolved over night in 0.5 ml TE buffer. Finally, the sample was treated with 20 g DNAse-free RNase for 1 h at RT. Typically, about 2 mg of chromosomal DNA were obtained.

4.2.5 Partial digestion of DNA

DNA was partially digested by a stepwise decrease in the concentration of restriction enzyme by serial dilution of enzyme with buffered DNA. The reaction mixtures were incubated for a fixed amount of time (5–30 min) at 37 C. Reactions were stopped by placing the tubes on ice water and immediately adding DNA loading buffer. In an initial experiment 15-l reaction mixtures containing 12 g DNA were used. After determination of the desired enzyme

concentration by analysis of the digests on agarose gels up to 150-l reactions were set up on the same day for subsequent preparative gel electrophoresis.

4.2.6 Transformation of *E. coli*

Electroporation of *E. coli* cells was performed with the Easyject (Eurogentec) electroporator. Electrocompetent cells were prepared as described by Dower (1988). Routinely, transformation efficiencies of $\sim 3 \cdot 10^7$ cfu / g pUC18 were achieved. Prior to electroporation ligation mixtures were desalted by precipitation with 2-butanol: The ligation mixture was brought to a final volume of 50 l with water, 450 l 2-butanol were added, followed by vortexing for 10 s. After centrifuging for 10 min at RT in a bench-top centrifuge and drying in a speed-vac the pellet was resuspended in 10 l of water.

$CaCl_2$ transformation was done according to Sambrook et al. (1989).

4.2.7 Construction of an *A. acidocaldarius* plasmid gene bank

Chromosomal DNA from *A. acidocaldarius* was partially digested (sec. 4.2.5) on a large scale in two batches of 120 g DNA in a volume of 150 l with 0.032 and 0.016 U *Sau*IIIA g^{-1} DNA. Restriction took place for 5 min at 37 C. After stopping the reaction, batches were mixed and $\frac{1}{3}$ of the mixture was applied to a gel for isolation of segments in the range from 2–7 kb. Finally, a ligation reaction was set up in a total volume of 20 l containing 200 ng *Bam*HI-digested and dephosphorylated pUC18 and 500 ng insert DNA. *E. coli* JM109 was transformed by electroporation with $\frac{1}{5}$ of the ligation mixture, yielding 20,000 transformants.

4.2.8 Screening of the plasmid gene bank and cloning of *celA*

Transformants were plated on large LB plates containing ampicillin and incubated over night at 37 C after which colonies were transferred to replica plates using Hybond N^+ membranes (Amersham). The new plates contained 1 mM IPTG. Again, colonies were allowed to grow over night. To prepare top agar 1.4 % oat spelt xylan was boiled for 10 s in 2 mM $MgSO_4$, 2.5 mM $CaCl_2$ and 80 mM β-alanine, pH 3.5 (60 C) and mixed with the same volume of 1.1 % Gelrite (Roth). 50 ml of the mixture was poured immediately onto each plate. After incubation over night at 57 C xylan-degrading activity was detected by flooding the plates with 0.1 % Congo red and destaining with 0.5 M NaCl according to Teather and Wood (1982).

The *celA* gene was subcloned into the expression vector pBAD/HisA by using a unique *Hind*III site 20 nt downstream of the stop codon (fig. 5.3) and introducing a *Sal*I restriction site 5' of the putative start codon (fig.

5.4) by site-directed mutagenesis (sec. 4.2.1). Primers used were
Sch260 5'ACGGCATGTCGACTCCCCTCCCACC3'
Sch261 5'CGGTGGGAGGGGGAGTCGACATGCCG3'
(mutated base underlined). Restriction produced a staggered end compatible with an *Xho*I site in the MCS of the vector. Ligation of the DNA fragment with the vector produced plasmid pFZ1 and resulted in a protein which had the following amino acids fused N-terminally to the CelA protein:

```
         10         20         30         39
MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSELD
```

4.2.9 Construction of an *A. acidocaldarius* λ Zap Express gene bank

Chromosomal DNA from *A. acidocaldarius* was isolated as described in sec. 4.2.4. After partial digestion of DNA (5 min at 37 C) in two batches of 105 g with 0.003 and 0.005 U g^{-1} *Sau*IIIA (sec. 4.2.5), DNA fragments ranging from 8–12 kb were isolated by preparative gel electrophoresis. 900 ng of insert DNA were ligated into 1 g of the Zap Express vector in a 5 l reaction mixture according to the manufacturer. For subsequent packaging the Gigapack III Gold packaging extract (Stratagene) was used. The gene bank was then titered and plated for blue-white colour selection using *E. coli* XL1-Blue MRF' as host strain according to the manufacturer's instructions. In order to assess the average insert size 10 randomly selected 'white' plaques containing inserts were subjected to single-clone excision of the pBK-CMV phagemid. Plasmids were isolated, cut with *Eco*RI / *Pst*I and the digests analyzed by gel electrophoresis.

4.2.10 Screening of the Zap Express gene bank and cloning of *celB*

Phages were plated on large NZY plates using *E. coli* XL1-Blue MRF' as host strain according to Stratagene. Plates were incubated over night at 37 C and cooled to 4 C. Then plaques were transferred to Hybond N^+ membranes. Replica plates were prepared by overlaying NZY plates with the host strain in LB top agar with 10 mM IPTG added. Immediately, the membranes were placed on the replica plates. After incubation for 6 h at 42 C plaques became visible and were screened for CMCase activity. This was done by overlaying the replica plates with 0.1 % CMC, 250 mM β-alanine, pH 3.5, 1 mM $MgSO_4$, 1.25 mM $CaCl_2$, 0.55 % Gelrite and incubating overnight at 57 C. The relatively high concentration of β-alanine buffer ensured a low pH of the top agar in order to select for acidophilic enzymes. Lysis zones around positive plaques were identified by staining with Congo red as described in sec. 4.2.8 prior to which the pH of the plates was adjusted by

flooding four times with 0.1 M Tris, pH 8. Plaques in lysis zones were picked and transferred to new plates and screened again. This was repeated, after which phagemids were derived and plated from positive plaques according to Stratagene using the ExAssist helper phage and the *E. coli* XLORL strain. The resulting plasmid harboring a 6.4 kb fragment was designated pKE25 (fig. 5.17A). In order to clone flanking regions of the original clone (pKE25) a terminal 565 bp *Xho*I / *Nco*I fragment was labelled and used in Southern hybridization with *Kpn*I digested chromosomal DNA from *A. acidocaldarius*. Thus a 5.7 kb *Kpn*I fragment could be identified and cloned into the vector pUC18, yielding pKE25a5 (fig. 5.17A).

4.2.11 Subcloning of the *celB* gene into pBAD/HisB

Plasmid pKE2201 was constructed by ligating a *PstI* / *Eco*RI fragment of pKE25 (fig. 5.17A, B) into the expression vector pBAD/HisB. The resulting ORF ($celB_{trunc}$) translates into a protein with the following residues fused to Gly 35 of the precursor:

```
         10         20         30       38
MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSRSA
```

Since the 3' region of the truncated ORF lacked a termination codon, a stop codon provided by the pBAD/HisB vector was used. This resulted in an extension of the protein by the sequence:

```
          11
PKNSKLGCFFG
```

carboxy-terminal of Asp 757. Plasmid pKE101, harboring the complete *celB* gene was constructed by fusing the inserts from pKE25a5 and pKE2201 via a unique *Kpn*I site in pBAD/HisB. Thus, recombinant full-length CelB has an N-terminus identical with $\mathrm{CelB_{trunc}}$, but is derived from the full-length ORF with the wild-type termination codon (fig. 5.17A).

4.3 Biochemical methods

4.3.1 Preparation of whole-cell extract

Whole-cell extract of *A. acidocaldarius* was prepared by resuspending 1.5 ml of a cell culture in 30 mM $MgSO_4$ at an OD_{650} of 24, followed by treatment with 3 mg lysozyme for 15 min at 37 C. For subsequent SDS-PAGE analysis the lysate was diluted 2× with loading buffer.

E. coli cells were lyzed by resuspending the cells at an OD_{650} of 12 in SDS-PAGE loading buffer and boiling for 5 min prior to electrophoresis.

4.3.2 Expression of the *celA* gene and purification of recombinant CelA

E. coli TOP10 (pFZ1) was grown in LB broth, containing ampicillin, to an OD_{650} of 0.5. Then, expression of *celA* was induced by arabinose addition (0.02 %) and growth continued for 4 h. Subsequently, cells were harvested, resuspended in buffer A (50 mM Na-phosphate buffer, pH 7, 300 mM NaCl, 0.1 mM phenylmethylsulfonyl fluoride (PMSF)) to an OD_{650} of 25 and disintegrated by sonication. After centrifugation at $130,000 \times g$ for 1 h at 4 C, the supernatant (cytosolic fraction) was incubated for 1 h at 60 C and a further 10 min at 70 C to denature host proteins and centrifuged again. The resulting supernatant (40 ml) was diluted 1 : 1 with buffer A, containing 20 mM imidazole, and subjected to Ni-NTA agarose chromatography (Qiagen) using 5 ml of resin. Binding was allowed to occur for 1 h at 4 C after which the protein-resin complex was transferred to a column (1.5 cm diameter) for washing (200 ml buffer A, containing 20 mM imidazole) and elution (35 ml buffer A, containing 100 mM imidazole). CelA-containing fractions were pooled, concentrated 5-fold with an Amicon concentrator equipped with a YM10 membrane, dialyzed overnight (12–16 kDa cut-off dialysis tubing, type 20 from Biomol, Germany) against buffer A, and stored at −80 C.

4.3.3 Purification of wild-type CelB

A. acidocaldarius cells were grown for three days on oat spelt xylan, reaching a final OD_{650} of 2 and harvested by centrifugation. Triton extraction was performed by resuspending the cells in the same volume of basal salts (sec. 4.1) with 0.05 % Triton X-100 added and incubating for 30 min at 57 C. Centrifuging for 15 min at $20,000 \times g$ yielded the Triton extract (supernatant). Routinely, 450 ml extract were adjusted to pH 6.5 by adding 10 mM BisTris buffer, and loaded onto a Q-Sepharose (Sigma) column (bed volume: 25 ml) equilibrated with 10 mM BisTris, pH 6.5, containing 0.94 mM $CaCl_2$, 2 mM $MgSO_4$, and 0.005 % Triton X-100 (buffer B). After washing with 150 ml buffer B, elution was performed with a NaCl gradient from 0 to 0.2 M in 500 ml of buffer B. CelB-containing fractions were collected, supplemented with β-alanine buffer, pH 3.5, to a final concentration of 40 mM and stored at −80 C.

4.3.4 Expression of the *celB* and $celB_{trunc}$ genes and purification of the recombinant proteins

E. coli strain TOP10 hosting either the plasmid pKE101 for production of full-length CelB or pKE2201 for $CelB_{trunc}$ was grown in LB broth, containing ampicillin, to an OD_{650} of 0.5. Expression of *celB* and $celB_{trunc}$ was induced by addition of 0.2 % and 0.02 % arabinose, respectively, and growth

continued for 4 h. Subsequently, cells were harvested and resuspended in buffer A (50 mM Na-phosphate, pH 7, 300 mM NaCl, 0.1 mM PMSF) to an OD_{650} of 25.

Purification of full-length CelB proceeded with subsequent disintegration of the cells by sonication for 5 min (Sonifier II, Branson) followed by centrifugation at $130,000 \times g$ for 1 h at 4 C. The resulting supernatant (2 ml) was mixed with 0.5 ml Ni-NTA agarose (Qiagen) and Tween 20 was added to a final concentration of 0.1 %. From hereon, Tween 20 and PMSF were present in all buffers. Binding took place for 30 min at 4 C at an imidazole concentration of 10 mM after which the matrix was transferred to a column and washed with 5 ml of buffer A containing 10 mM imidazole. Elution of bound protein was performed by raising the imidazole concentration stepwise from 25 to 200 mM. CelB-containing peak fractions were pooled and dialyzed overnight (dialysis tubing type 20, 12–16 kDa cut-off, Biomol, Germany) against buffer D (50 mM β-alanine, pH 3.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$).

$CelB_{trunc}$ was purified by disrupting the resuspended cells in a French press at 18,000 psi. After centrifugation 50 ml of the resulting supernatant were diluted 1 : 1 with buffer A and incubated with 5 ml Ni-NTA agarose for 30 min in the presence of 10 mM imidazole. Then, the resin was transferred to a column, washed with 50 ml buffer A, containing 20 mM imidazole and protein was eluted with 65 ml buffer A, containing 50 mM imidazole. Peak fractions were pooled, concentrated 10-fold with an Amicon concentrator (YM30 membrane) and dialyzed overnight against buffer D.

4.3.5 Enzyme assays

Reducing sugar assay

Enzymatic activity against polysaccharides was quantified by measuring release of reducing sugars according to Somogyi (1951). 1 Unit is defined as the amount of enzyme releasing 1 mol reducing equivalents or 1 mol *p*-nitrophenol (pNP) per minute.

Polysaccharide stock solutions were prepared in different ways. Oat spelt xylan (Roth, Karlsruhe, Germany) was solubilized by stirring for 1 h at pH 10. Subsequent centrifugation for 15 min at $20,000 \times g$ resulted in a clear supernatant at a concentration of 1–2 %, determined by drying in a speed-vac and weighing the pellet. Phosphoric acid-swollen cellulose was prepared according to Wood (1988) and Wang *et al.* (2001), whereby Avicel PH-101 (Fluka) was stirred for 1 h at 4 C in 85 % H_3PO_4, centrifuged and washed three times with water and resuspended in buffer C (50 mM β-alanine, pH 3.5 (70 C), 0.94 mM $CaCl_2$, 2 mM $MgSO_4$). Birchwood xylan (2 % stock, Roth), arabinan (10 % stock, Megazyme, Ireland), laminarin (Sigma) and CMC (2 % stock, Serva Feinbiochemica) were dissolved by stirring in water

at RT. Starch (2 % stock, Sigma) and linear arabinan (6.7 % stock) from sugar beet (Megazyme) were boiled briefly. Lichenan (*Cetraria islandica*, Sigma) was dissolved at a concentration of 1 % at 80 C.

Equal volumes of substrate and enzyme solution were mixed and incubated. Following incubation, samples were cooled briefly on ice and 0.2 ml of the reaction mixture was mixed with the same volume of reagent D (see below), boiled for 20 min in a water bath and cooled. Subsequently, reagent C was added 1 : 1. After the reaction had proceeded for 20 min at RT with repeated vortexing, 0.5 ml were centrifuged for 10 min in a bench top centrifuge and the absorbance of the supernatant was measured at 520 nm in an ELISA plate reader. Reducing sugar content was calculated using a standard curve with 0–1 mM monosaccharide solution (glucose, xylose or arabinose) instead of enzyme solution. Reducing equivalents due to enzymatic activity were calculated by comparison with a reaction mixture whereby enzyme and substrate solutions were incubated seperately prior to estimation of reducing sugar content.

Under standard conditions CelA activity was assayed in buffer E (50 mM Na acetate, 200 mM NaCl, 10 mM $CaCl_2$, pH 5.5) at 60 C for 9 min, using 0.5 % CMC as substrate at an enzyme concentration of 0.18 mg ml^{-1}, while CelB was employed at a protein concentration of 1.3 g ml^{-1} in buffer C (50 mM β-alanine, pH 3.5 (70 C), 0.94 mM $CaCl_2$, 2 mM $MgSO_4$) with 0.25 % CMC and incubated for 1 h at 70 C.

The composition of the reagents is as follows:

Reagent A:

2.5	%	Na_2CO_3
2	%	$NaHCO_3$
3.35	%	K-Na-tartrate·$4H_2O$
20	%	Na_2SO_4

Reagent B: 15 % $CuSO_4 \cdot 5H_2O$

Reagent C:
6.25 g ammonium molybdate are dissolved in 112 ml of water to which 5.3 ml of concentrated H_2SO_4 have been added. 0.75 g Na_2H arsenate·$7H_2O$ are dissolved in 6.3 ml water and mixed with the molybdate solution with continuous stirring using a Pasteur pipette. The final volume is adjusted to 125 ml and the reagent incubated for 30 min at 55 C.

Reagent D: 25 ml reagent A + 1 ml reagent B, mixed directly before use.

Assay of *p*-nitrophenyl-glycoside hydrolysis

Hydrolysis of the chromogenic substrates *p*-nitrophenyl-α-L-arabinofuranoside, *p*-nitrophenyl-β-D-glucoside, *p*-nitrophenyl-β-D-cellobioside, *p*-nitrophenyl-β-D-cellotrioside and *p*-nitrophenyl-β-D-cellotetraoside (pNPAraf, pNPG, pNPC, pNPG3 and pNPG4, respectively, Sigma) was assayed by following the release of *p*-nitrophenol (pNP). The reaction was started by adding enzyme to preheated substrate in buffer in Eppendorf tubes, vortexing and incubating in a thermoblock. After the appropriate amount of time, 33 l were removed and mixed with 167 l of 0.125 M NaOH in an ELISA plate. pNP was then quantified by measuring the absorbance at 410 nm in an ELISA plate reader and comparing with a standard curve using pNP (Deshpande et al. (1984)).

Under standard conditions activity of CelA was assayed at enzyme concentrations between $1.8 \cdot 10^{-3}$ and 0.18 mg ml^{-1} in buffer E (50 mM Na acetate, 200 mM NaCl, 10 mM $CaCl_2$, pH 5.5) for up to 20 minutes at 60 C, while CelB was incubated for 1 h at 70 C in buffer C (50 mM β-alanine, pH 3.5 (70 C), 0.94 mM $CaCl_2$, 2 mM $MgSO_4$) at a protein concentration of 1.3 g ml^{-1} with 10 mM pNPAraf.

4.3.6 Determination of pH and temperature stability

Temperature stability of CelA was determined by incubating concentrated protein (0.3 mg ml^{-1}) at the indicated temperatures in buffer containing 10 mM $CaCl_2$ or 2 mM EDTA or no supplements. After heat treatment the protein was diluted 180-fold in buffer containing 10 mM $CaCl_2$. Residual activity was then measured in the dilution buffer under standard conditions, i.e. at 60 C and compared with the activity of protein which had been incubated on ice instead of heat treated and to which $CaCl_2$ had been added. Similarly, pH stability was assayed by incubating concentrated protein (0.075 mg ml^{-1}) overnight at 4 C at different pH values in 0.2 M buffer solution. Following a 45-fold dilution in buffer E (50 mM Na acetate, 200 mM NaCl, 10 mM $CaCl_2$, pH 5.5), residual activity was measured under standard conditions.

In the case of CelB pH stability was determined by incubating concentrated $CelB_{trunc}$ (25 g ml^{-1}) in 75 mM of the indicated buffers, supplemented with 10 mM $CaCl_2$ and 10 mM $MgCl_2$. After incubation overnight at 4 C, the sample was diluted 40-fold in buffer D (50 mM β-alanine, pH 3.5 (RT), 10 mM $CaCl_2$, 10 mM $MgCl_2$) and activity was assayed under standard conditions.

4.4 Crystallization of CelA

Crystallization was performed using the hanging-drop vapour-diffusion method in Linbro plates. 1 ml reservoir solution was employed. Drops consisted of 2 l protein solution and 2 l reservoir. After a solubility screen according to Stura et al. (1992) initial crystallization trials were performed with Crystal screens I and II (Hampton research) with CelA in the original phosphate buffer (sec. 4.3.2). Because of crystallization of phosphate the original buffer was removed from subsequent batches by concentration and dilution with 50 mM Tris-Cl, 0.1 mM PMSF, pH 7.0. Later, CelA was immediately dialyzed against Tris buffer (10 mM Tris-Cl, 0.1 mM PMSF, pH 7.0) upon purification.

When concentrating CelA above 0.5 mg ml^{-1} a Milipore Ultrafree-4, 10,000 kDa cut-off, polysulfonate based membrane was used as CelA was found to precipitate on cellulose based membranes.

The Hampton Research Additive screens 1, 2 and 3 were employed as suggested by the manufacturer. The protein dye Izit (Hampton Research) was used by adding 1 l of 10-fold diluted dye to the drop.

4.5 Analytical methods

4.5.1 Protein concentration assay

Protein concentration was determined with the Micro BCA Protein Assay Reagent Kit (Pierce) using bovine serum albumins as standard (Pierce).

4.5.2 Amino acid sequence analysis

Protein samples (40-fold concentrated Triton extract or purified wild-type CelB) were subjected to SDS-PAGE using large gels containing 7.5 % acrylamide and stained with Serva Blue R (sec. 4.5.3), after which CelB was excised. Cyan bromide treatment and N-terminal amino acid sequencing were kindly performed by R. Schmid (University of Osnabrck) as described (Schneider et al. (1994), Hner zu Bentrup et al. (1994)).

4.5.3 SDS-PAGE and staining procedures

SDS-PAGE was carried out according to Laemmli (1970). Samples were boiled for 5 min in SDS reducing sample buffer prior to application to the gel. Routinely, small gels with a separating gel length of 6 cm were run at 150 V using a Mini Protean II apparatus (Biorad). Large gels (12 cm separating gel) were run in a V16 tank (Life Technologies). For estimation of molecular mass the Broad Range Protein Standard (Life Technologies) was employed. The acrylamide concentration in the separating gel was between

7.5 and 10 %. Proteins were stained either with Serva Blue R or silver stained as described in tab. 4.3 (Blum et al. (1987)).

4.5.4 Antiserum production and Western blotting

Polyclonal antiserum production was performed by Biogenes, Germany. The rabbits were immunized either with purified wild-type CelB or purified recombinant CelA.

Immunoblot analyses were performed by transferring proteins from small SDS gels (sec. 4.5.3) onto nitrocellulose membranes using a 'Trans-Blot semi-dry' apparatus (BioRad, Towbin et al. (1979)). Blotting was performed for 25 min at 20 V in 25 mM Tris, 192 mM glycine, 20 % methanol, pH 8.3. Subsequently proteins were stained with 0.2 % Ponceau S in 3 % trichloroacetic acid. After brief washing with TNT buffer (20 mM Tris, 50 mM NaCl, 0.05 % (v/v) Tween 20, pH 7.5), lanes and molecular weight standard were marked and the membrane was completely destained in TNT buffer. Non-specific binding sites on the membrane were then blocked in 10 ml of 5 % skimmed milk powder in TNT buffer (blocking buffer). Subsequently, polyclonal antiserum was added and incubated over night at 4 C with the membrane. After washing (3x for 10 min with TNT buffer) binding of peroxidase-conjugated donkey anti-rabbit immunoglobulins to the antigen-antibody complexes took place for 40 min at 37 C in 10 ml blocking buffer. Then the membrane was washed again. Complexes were visualized with enhanced chemiluminescence (Luminol, NEN, USA) and exposure to Hyperfilm (Amersham-Buchler, Germany).

4.5.5 Zymogram analysis

For activity staining SDS gels containing either 0.2 % CMC or 0.2 % oat spelt xylan, solubilized as described in sec. 4.3.5, were used. Electrophoresis took place at 100 V after which gels were treated with buffer C (50 mM β-alanine, pH 3.5 (70 C), 0.94 mM $CaCl_2$, 2 mM $MgSO_4$) according to Chen and Buller (1995). One small gel was washed twice at room temperature in 250 ml buffer C containing 25 % isopropanol followed by a wash in the same volume of buffer C without additional isopropanol. Finally, enyme activity was induced by submerging the gel in 250 ml of buffer C at 57 C for 1 h. After neutralizing the pH of the gel by washing for 30 min in 500 ml of 50 mM Tris-Cl, pH 8, the gel was stained for 10 minutes with 0.1 % Congo red. Activity bands, were made visible by destaining with several washes of 1 M NaCl.

Step	Solution	Time
fixation	100 ml 10 % ethanol	2×5 min
wash	100 ml water	3×3 min
pretreatment	100 ml 0.02 % Na-thiosulfate	1 min
wash	100 ml water	3×20 s
impregnation	100 ml 0.1 % $AgNO_3$, 23 l 37 % formaldehyde	30 min
developement	100 ml 6 % Na_2CO_3, 0.4 mg Na-thiosulfate, 57 l 37 % formaldehlyde	until bands are visible
stop	100 ml 0.05 M EDTA	10 min

Table 4.3: Silver staining procedure for SDS gels. Amounts are given for one small gel. Adapted from Blum et al. (1987).

4.5.6 Thin-layer chromatography

Substrates (sec. 4.3.5, 4.3.5) were digested in the apropriate buffers (standard conditions, sec. 4.3.5). Subsequently, aliquots (1 l) were spotted on a silica 60 TLC plate (Merck), which was developed in ethyl acetate / acetic acid / water (2 : 1 : 1) for three hours according to Hrmova and Fincher (1993). After drying, sugars were visualized by short immersion of the plate in orcinol ferric chloride reagent (Sigma) and heating to 100 C.

4.5.7 Analysis of metal binding

EDXRF (energy dispersive X-ray fluorescence spectrometry) analysis was performed after two dialyses over night against double-distilled water (Millipore, 1 : 8300) in a plastic beaker. 85 g of CelA were applied in each of two measurements. Excitation of the sample was achieved using a high power MoKα beam monochromatized by a highly oriented pyrolytic graphite crystal. Detection of the resulting fluorescence spectrum was carried out with a Si(Li) detector and subsequently analyzed using the X-Spec32 computer code. The metal content of the protein was estimated relative to its sulfur content (Laursen (1985), Laursen et al. (1998)).

4.6 Computer aided analyses

DNA and protein sequences were analyzed using DNASIS (Hitachi). The hydropathy plot was obtained using the algorithm of Kyte and Doolittle (1982) with a window size of 50. N-terminal signal peptides were detected using SignalP (Nielsen et al. (1997)). Database searches were conducted with BLASTP 2.2.5 and PSI-BLAST at NCBI (Altschul et al. (1997)). Internal sequence similarities and local alignments between two sequences

were analysed using PLALIGN 2.1 (Huang and Miller (1981)). ClustalX (Thompson et al. (1997)) was used for alignments and construction of phylogenetic trees with the neighbour-joining method. Figures were drawn with Genedoc (Nicholas et al. (1997)) and Treeview (Page (1996)). The Jpred server (Cuff et al. (1998)) and the 3D-PSSM server (Kelly et al. (2000)) were used for secondary structure predictions. Structural alignments of known 3D structures were made and displayed with Swiss Pdb Viewer 3.7.

Classification of glycoside hydrolases was looked up on the CAZy server[1] (Coutinho and Henrissat (1999)).

4.7 Nucleotide sequence accession numbers

The complete nucleotide sequences of *celA* and *celB* have been deposited in the EMBL Nucleotide Sequence Database under the accession numbers AJ308623 and AJ551527.

[1] http://afmb.cnrs-mrs.fr/CAZY/GH.html

5 Results

5.1 Gene cloning and characterization of CelA

5.1.1 Construction and screening of an *A. acidocaldarius* plasmid gene bank

In order to isolate genes coding for xylan-degrading enzymes a genomic library of *A. acidocaldarius* was constructed and screened for clones exhibiting xylan-degrading activity on plates. pUC18 was used as a vector in an *E. coli* JM109 host strain. A gene bank with overlapping inserts was obtained by subjecting chromosomal DNA to partial digestion with *Sau*IIIA and isolating fragments in the range from 2–7 kb. Fig. 5.1 shows the initial experiment used to determine the required concentration of *Sau*IIIA in the digests. Screening of the resultant transformants with the substrate overlay method using oat spelt xylan yielded one xylanase-positive colony out of all the 20,000 transformants obtained. Fig. 5.2 shows detection of xylanase activity in this clone. The respective plasmid pKE1 contained a 3.5 kb insert.

5.1.2 Identification of the *celA* gene

Because initial sequencing of the 5' and 3' region of the pKE1 insert revealed no homology to glycosyl hydrolases, pKE1 was analyzed by restriction mapping. This revealed a 3.5 kb insert. Four subclones (pKE2–5) were constructed and tested for xylanase activity with the substrate overlay method. Only one of the subclones (pKE2) gave rise to xylanase activity (fig. 5.3).

Plasmid pKE2 contained a 2.8 kb *Eco*RI / *Hind*III fragment of the pKE1 insert. Nucleotide sequencing revealed a complete ORF encoding a putative protein of 537 amino acids with a molecular mass of 59 kDa. A data base search indicated high sequence similarity to cellulases, so the putative protein was named CelA, consistent with subsequent experiments on substrate specificity. A second only partially cloned ORF upstream of the *celA* gene encodes a protein with sequence similarity to various quinolinate synthetases, aligning best (45 % identical residues) with the NadA protein of *Mycobacterium tuberculosis*. A potential transcriptional terminator for

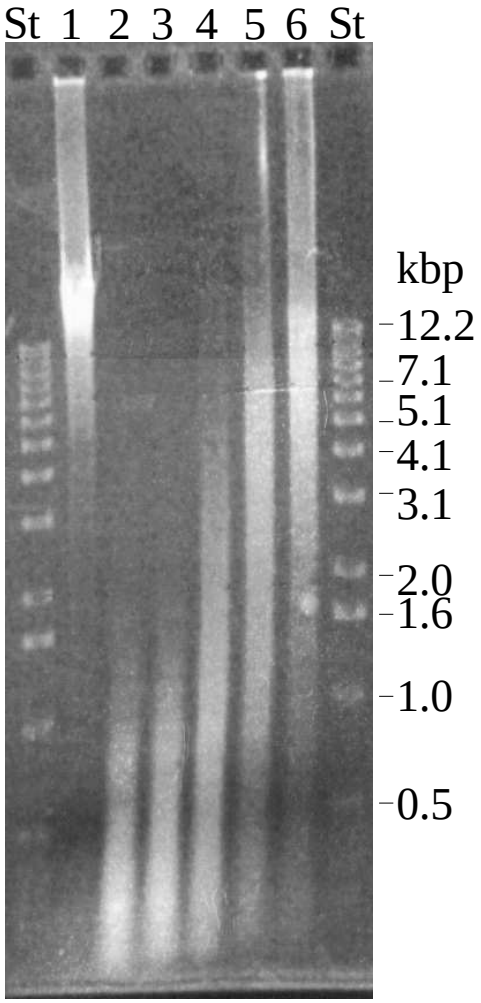


Figure 5.1: Agarose gel electrophoresis of *Sau*IIIA partial digests of *A. acidocaldarius* chromosomal DNA. *St*, DNA size standard; *lane 1*, undigested DNA; *2–6*, DNA digested with decreasing concentrations of restriction enzyme: *2*, 0.25 U g^{-1} DNA ; *3*, 0.125 U g^{-1} DNA; *4*, 0.063 U g^{-1} DNA; *5*, 0.032 U g^{-1} DNA; *6*, 0.016 U g^{-1} DNA; *1–6*, 12 g DNA per lane. Restriction took place for 5 min. Reactions from lanes 5 and 6 were upscaled and fragments isolated in the range from 2–7 kb for the construction of the plasmid gene bank.

Figure 5.2: Clone displaying xylanase activity (*E. coli* JM109⟨pKE1⟩) and the same strain harboring the vector plasmid (pUC18) overlayed with xylan. After growth on LB containing ampicillin the colonies were overlayed with xylan at pH 3.5 and incubated over night at 57 C. Subsequently xylan was stained with Congo red.

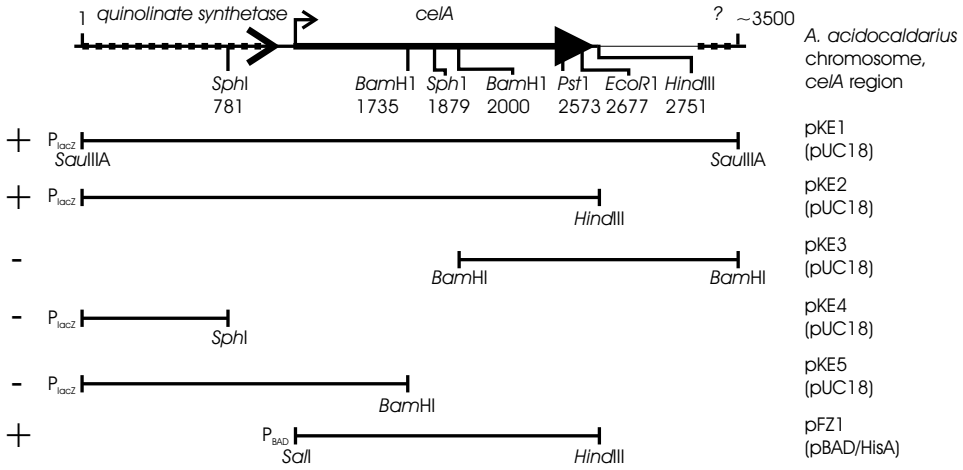


Figure 5.3: Overview of the *celA* region and cloning strategy. Shown is the *celA* region of the *A. acidocaldarius* chromosome (*top line*). *Numbers* indicate nucleotide positions relative to the 5' *Sau*IIIA site of the original clone (pKE1). ORFs are represented by *arrows* in the direction of transcription. The *celA* ORF is made up of nt 1121–2731, the putative catalytic domain corresponds to nt 1367–2731. *Dashed arrows* show ORFs neighbouring *celB* with putative assignments. The *crooked arrow* indicates the *celA* promotor. *Thin line* indicates region for which no sequence information is available. Restriction sites relevant to the cloning strategy are given. Below inserts of the constructed plasmids are drawn in relation to the *celA* region with the restriction sites used for excision of the insert prior to ligation in the host vector (in *parentheses*). P_{BAD} and P_{lacZ} indicate the position of the host vector promotor if known. On the far left is indicated whether xylanase activity was detected with the substrate overlay method (+) or not (-).

the putative quinolinate synthetase gene with a free energy value of −62 kcal·mol^{-1} was identified between nucleotides 88 to 170 (fig. 5.4). Downstream of *celA* an incomplete ORF is located that translates into a protein with no match in the data base.

A possible start codon (ATG) was found in the 5' region of the *celA* gene at a position to which a putative ribosome binding site and −10 promotor site could be assigned (fig. 5.4). However, no −35 sequence was identified. No N-terminal signal peptide was detected using the program SignalP.

5.1.3 Sequence analysis of CelA

Sequence alignment of CelA with the entries in the NCBI databases using BLASTP revealed homology to members of glycoside hydrolase family 9 (Henrissat (1991)), containing β-1,4-endoglucanases. Thus and in accordance with a recent proposal an alias protein name would be Cel9A.

Enzymes belonging to family 9 display a common catalytic domain with an $(\alpha/\alpha)_6$ twisted α-barrel motif (Beguin and Alzari (1998), fig. 5.5). Highest sequence identity (33 % identity, 47 % homology) was displayed towards a putative endo-1,4-β-D-glucanase from *Xanthomonas campestris pv. campestris* and a cellodextrinase (Ced1) from *Butyrivibrio fibrisolvens* H17c (34 % identity, 50 % similarity, Berger et al. (1990)). No hits belonging to other glycoside hydrolase families were produced. Fig. 5.6B shows a sequence comparison between the catalytic domains of CelA, Ced1 and two other family 9 cellulases, CelD and E4, for which 3D structures are available. CelD from *Clostridium thermocellum* displays an N-terminal Ig-like domain followed by the catalytic domain (Juy et al. (1992), reviewed in Beguin and Alzari (1998)) and belongs to subgroup E1. In the endoglucanase E4 from *Thermomonospora fusca*, belonging to subgroup E2, the catalytic domain is located at the N-terminus followed by two CBDs, one of which is included in the crystal structure (Sakon et al. (1997)). Sequence identities in the catalytic domain were 20 % between CelA and E4, and 24 % between CelA and CelD.

Compared to a structural alignment (fig. 5.5) between the active sites of E4 and CelD, the sequence of CelA showed conservation not only of the active site carboxylates (Asp 146 and Glu 515 in CelA), but also of some aromatic and basic residues which are part of conserved residue blocks. These are summarized in table 5.1 and highlighted in figs. 5.5 and 5.6. As can be seen from fig. 5.5 no conserved aromatics can be detected in the CelA sequence for subsites distal from the scissile bond.

PSI-BLAST aided secondary structure prediction using the Jpred server on the N-terminal portion of CelA showed an all β-sheet structure prediction for this domain (fig. 5.6A). A fold recognition search with the 3D-PSSM server had the Ig-like domain of CelD as the top scoring fold, although with a low score, with several other Ig domains among the top hits. From this

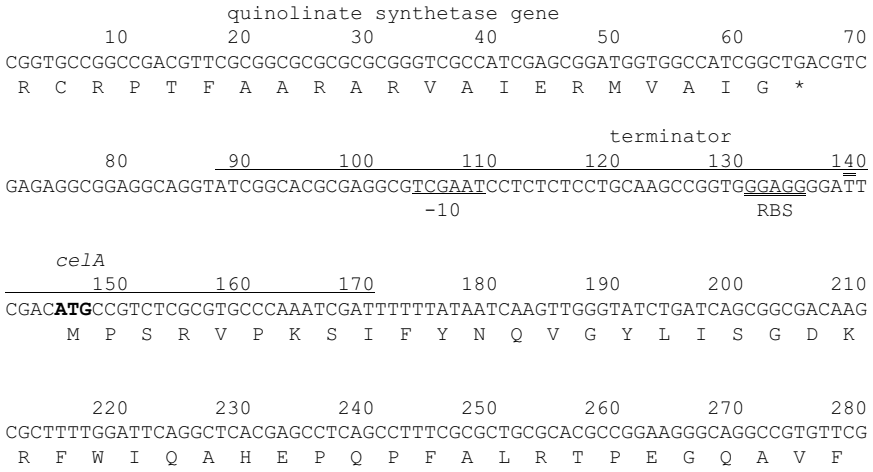


Figure 5.4: Sequence analysis of the 5 region of the *celA* gene. Shown are the nucleotide sequence and the corresponding amino acids for the putative quinolinate synthetase gene and *celA*. The putative ribosome binding site (RBS, *double-underlined*), the terminator (*overlined*), and the −10 region (*underlined*) are indicated. Possible start codon is in *boldface*. The T nt changed to a G for cloning of *celA* into pBAD/HisA, yielding pFZ1, is *double-overlined*.

CelA	CelD	E4	Function
Asp 146	Asp 201	Asp 104	general base
Tyr 300	Tyr 354	Tyr 252	subsite −1
Trp 401	Trp 457	Trp 359	subsite −2
His 461	His 516	His 422	subsite +1/+2
Glu 515	Glu 555	Glu 470	general acid
Trp 520	Trp 560	Tyr 475	subsite −1

Table 5.1: Positions of conserved residue blocks in CelA (*A. acidocaldarius*), CelD (*C. thermocellum*) and E4 (*T. fusca*). The subsites associated with the aromatic platforms and the role of the catalytic carboxylates are given. See also figs. 5.5 and 5.6.

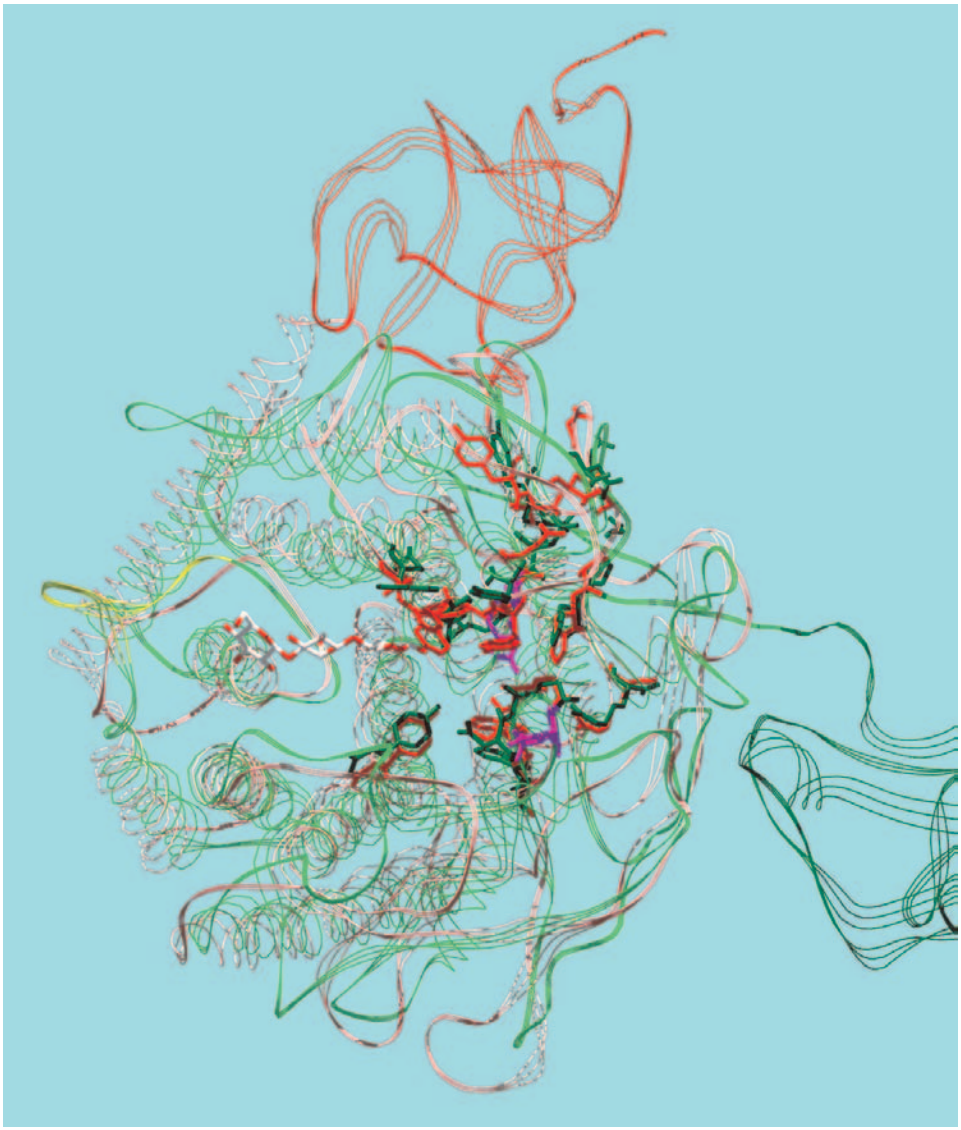
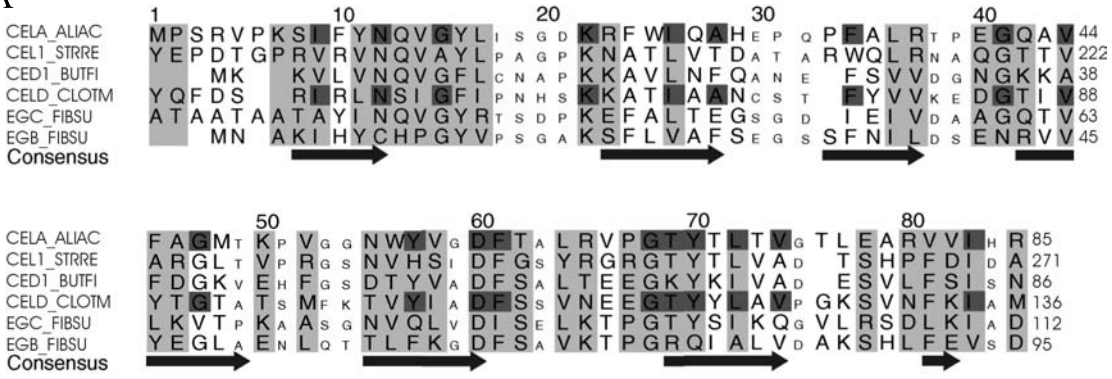


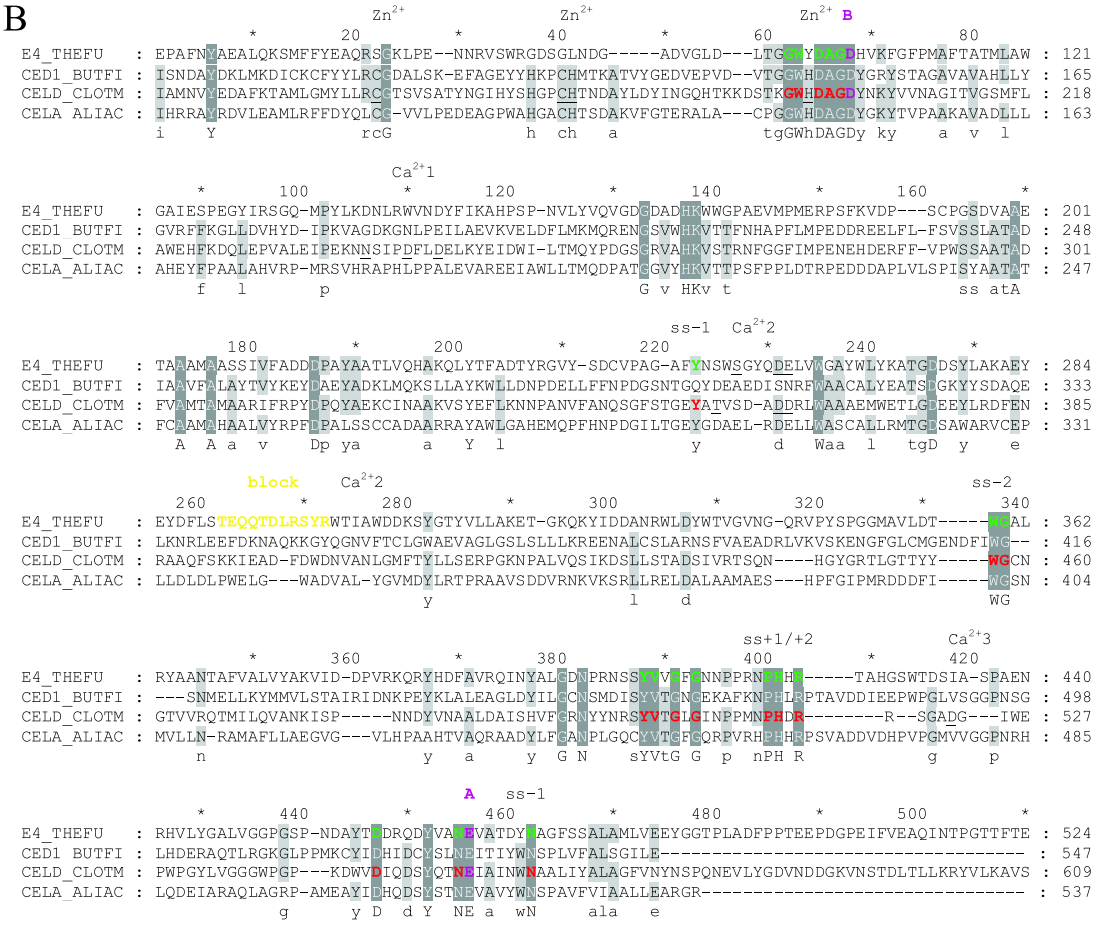
Figure 5.5: Structural alignment (ribbon diagram) of catalytic domains of the two known GH family 9 3D structures. E4-68, a proteolytic fragment of E4 from *T. fusca*, is shown in *green* with the partially visible C-terminal CBD in *dark green* (Sakon et al. (1997), PDB code 1js4). The block in E4 not present in CelD to the non-reducing side of the −4 subsite is in *yellow*. CelD from *C. thermocellum* is shown in *red* with the N-terminal Ig-like domain in *dark red* (Juy et al. (1992), PDB code 1clc). Amino acids belonging to conserved residue blocks (coloured accordingly in fig. 5.6) are shown with *side chains*. Catalytic residues are in *violet*. Glucose units (*CPK wireframe*) occupy subsites −4, −3 and −2.

Figure 5.6: Sequence alignment of CelA (CELA_ALIAC) with representatives of family 9 glycoside hydrolases. The proteins considered are Cel1 from *Streptomyces reticuli* (CEL1_STRRE), endoglucanase E4 from *T. fusca* (E4_THEFU, pdb code 1js4a), Ced1 from *Butyrivibrio fibrisolvens* H17c (CED1_BUTFI), CelD from *C. thermocellum* (CELD_CLOTM, pdb code 1clc), EGC from *Fibrobacter succinogenes* BL2 (EGC_FIBSU) and EGB from *F. succinogenes* S85 (EGB_FIBSU). **(A) Alignment of Ig-like domains made with Alscript (Barton (1993)).** *Light grey shading* shows conserved similar residues in all sequences. Identical residues between CELA_ALIAC and CELD_CLOTM are in *grey*. Consensus shows regions of predicted β-strands from the Jpred server. **(B) Alignment of catalytic domains.** The alignment was carried out using the program ClustalX and edited by eye. The CelD and E4 sequences continue beyond the shown alignment. Identical residues in three and all four proteins are high-lighted in *light gray* and *gray*, respectively. Catalytic acid and base (Beguin and Alzari (1998), Tomme et al. (1991), Tomme et al. (1992)) are marked by *A* and *B*, respectively. The positions of the zinc binding site (Zn^{2+}) in CelD and three calcium binding sites (Ca^{2+}) are indicated for CelD and E4 according to Juy et al. (1992) and Sakon et al. (1997). Only $Ca^{2+}2$ is present in E4. Amino acids which show side chain interactions with bound cations are *underlined*. Amino acids belonging to conserved residue blocks are coloured in accordance with fig. 5.5 and assigned to a subsite (*ss*). The steric block in E4 at the non-reducing end of the active site cleft is indicated (*block*) in *yellow*.

A


CELA_ALIAC
CEL1_STRRE
CED1_BUTFI
CELD_CLOTM
EGC_FIBSU
EGB_FIBSU
Consensus

B


E4_THEFU
CED1_BUTFI
CELD_CLOTM
CELA_ALIAC
Zn2+
Ca2+1
ss-1
Ca2+2
block
ss-2
ss+1/+2
Ca2+3
A
B

analysis, CelA is likely to possess the Ig-like domain and thus, can be placed in the E1 group of family 9. This conclusion is also consistant with the lack of a CBD. Thus, the C-terminal catalytic domain of CelA is comprised of residues Ile 83–Arg 537 while the preceding N-terminal 82 amino acids constitute an Ig-like domain. Only clones harbouring the complete *celA* ORF gave rise to xylanase activity, while truncated ORFs did not. This is not surprising, since none of the truncated proteins contained the complete catalytic domain (fig. 5.3).

5.1.4 Purification of recombinant CelA

The *celA* gene was subcloned into an expression vector, placing the gene under control of an arabinose dependent promotor and adding an N-terminal His_6-tag to the CelA protein.

Thus, CelA was purified as an N-terminal His_6-fusion protein from the cytosolic fraction of *E. coli* strain TOP10 (pFZ1) by Ni-NTA chromatography (fig. 5.7A). Typically, from a 1 l culture 80 ml of cytosol were obtained with a specific activity of 16 U mg^{-1} on CMC. Purified CelA displayed a specific activity of 73 U mg^{-1} on CMC with a total activity of 511 U, corresponding to a protein yield of 7 mg. The SDS gel of the purified and concentrated preparation (fig. 5.7A, lane 3) revealed a minor contaminant with a molecular mass of 30 kDa, which is likely to originate from the *E. coli* host. In order to assure that the enzymatic activities displayed by the preparation are not due to this protein, cells of *E. coli* TOP10, harboring the vector plasmid pBAD/HisA, were subjected to the purification protocol established for CelA. Samples from the final step indeed contained a 30 kDa protein (fig. 5.7A, lane 4), but did not display any activity on pNPC or CMC.

5.1.5 Expression of *celA* in *A. acidocaldarius*

Expression of the *celA* gene was demonstrated in *A. acidocaldarius* by Western blot analysis, using polyclonal antibodies directed against the purified recombinant protein. CelA was detected when cells were grown with CMC or oat spelt xylan as sole carbon and energy source. Low levels of expression were also found after cells were grown on glucose, cellobiose, starch and birchwood xylan(fig. 5.7B).

5.1.6 pH and temperature dependency of CelA

At 60 C CelA displayed a rather broad pH optimum around 5.5, with still 50 % activity at pH 4.7 or pH 7 (fig. 5.8 A). When stored overnight at 4 C and at pH values between 5 and 9, no loss of activity was observed. However outside of these values stability dropped sharply (fig. 5.8 B). Thus, the pH

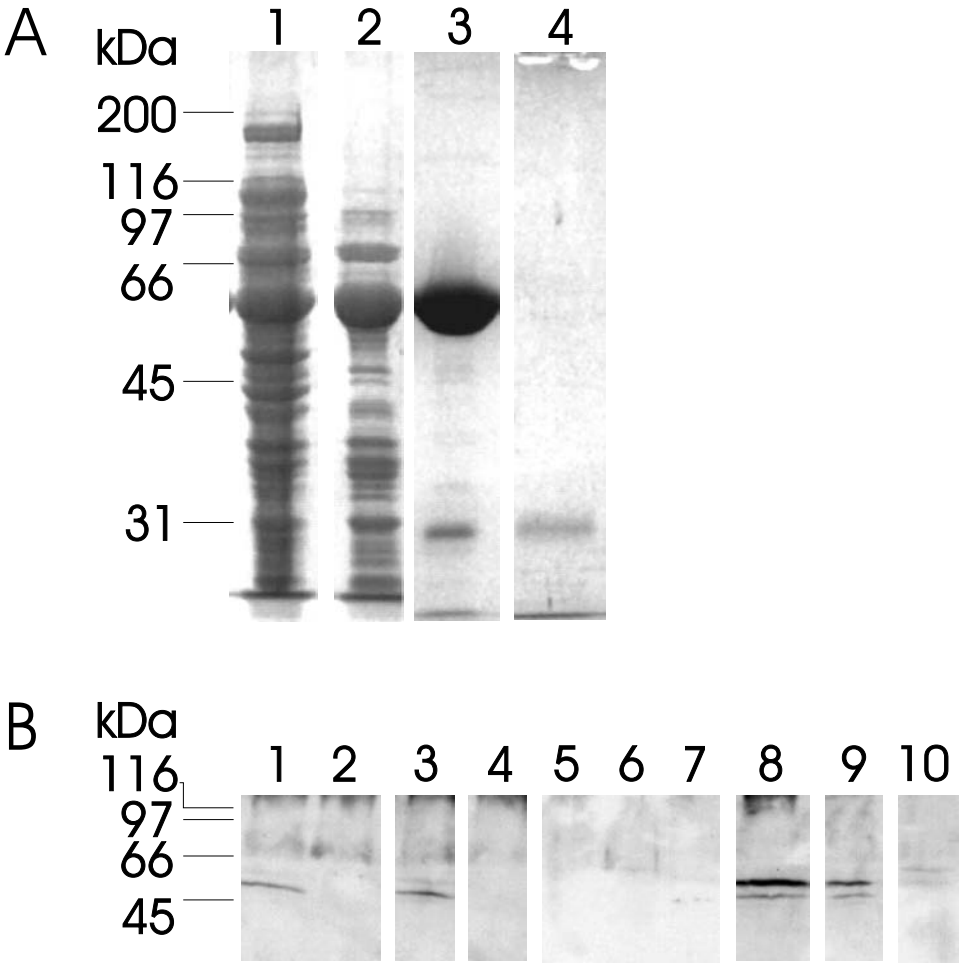


Figure 5.7: (A) Overproduction and purification of CelA. Protein samples from each step were subjected to SDS-PAGE and subsequently stained with Coomassie brilliant blue R250. *Lane 1*, Cytosolic fraction of induced cells of *E. coli* TOP10 (pFZ1) (39 g); *2*, supernatant after heat denaturation (13 g); *3*, pooled and concentrated CelA-containing fractions from Ni-NTA chromatography (8 g); *4*, as in *3*, but obtained from *E. coli* TOP10 (pBAD/hisA) (control). **(B) Detection of CelA in whole-cell extracts of *A. acidocaldarius*.** Wild-type *A. acidocaldarius* was grown in minimal salt medium supplemented with the following carbon sources: *Lane 1*, glycerin; *2*, glucose; *3*, xylose; *4*, maltose; *5*, cellobiose; *6*, starch; *7*, arabinan; *8*, CMC; *9*, oat spelt xylan; *10*, birchwood xylan. Proteins from whole-cell extracts (25 l per lane) were separated by SDS-PAGE and subjected to immunoblot analysis.

characteristics of CelA did not match with the optimal pH of 3.5 for growth of *A. acidocaldarius*.

At pH 5.5 the temperature optimum was determined to be 70 C, retaining 50 % activity at 50 and 80 C, respectively (fig. 5.9 A). The enzyme was relatively stable at 60 C with little loss of activity after one hour even without addition of $CaCl_2$. However, at 75 C the enzyme had a half life of only 30 min. Under these conditions the presence of 10 mM $CaCl_2$ resulted in an increase of residual activity from 50 % to 70 %, whereas the presence of 2 mM EDTA resulted in total loss of residual activity (fig. 5.9 B).

5.1.7 Metal binding by CelA

X-ray fluorescence analysis (EDXRF) of CelA after extensive dialysis against water revealed an S : Ca : Ni : Zn ratio of 28 : 1.9 : 0.02 : 0.92, most consistent with the presence of two tight binding sites for calcium and one for zinc in each CelA molecule. According to this analysis essentially no nickel remained attached to the His-tag after dialysis.

5.1.8 Substrate specificity of CelA

Initially identified as a xylan-degrading enzyme, purified CelA was confirmed to display low activity on oat spelt xylan. However, high activities were observed for β-1,4-linked glucans, such as lichenan and CMC (tab. 5.2). No activity was found on starch, laminarin or birchwood xylan. TLC analysis of products derived from CMC digestion revealed only little glucose and cellobiose even after 2 hours of digestion (fig. 5.10A). Nonetheless the viscosity of 0.5 % CMC was considerably decreased on incubation with CelA (55 g ml^{-1}), as the amount of time required for 1 ml of the digest to travel down a vertical glass pipette decreased from 18 ± 2.5 s to 10 ± 1 s after 1 h of digestion at 60 C. A time course of lichenan digestion (fig. 5.10B) showed that

Substrate	Specific activity (U mg^{-1})
CMC	73
Lichenan	109
Oat spelt xylan	6
pNPG	0
pNPC	2.1
pNPG3	0
pNPG4	0.3

Table 5.2: Specific activity of CelA (0.18 mg ml^{-1}) towards aryl glycosides (10 mM) and polysaccharides (0.5 %). Specific activity determined by assay of pNP or reducing sugar release, respectively. No activity was found on starch, laminarin or birchwood xylan.

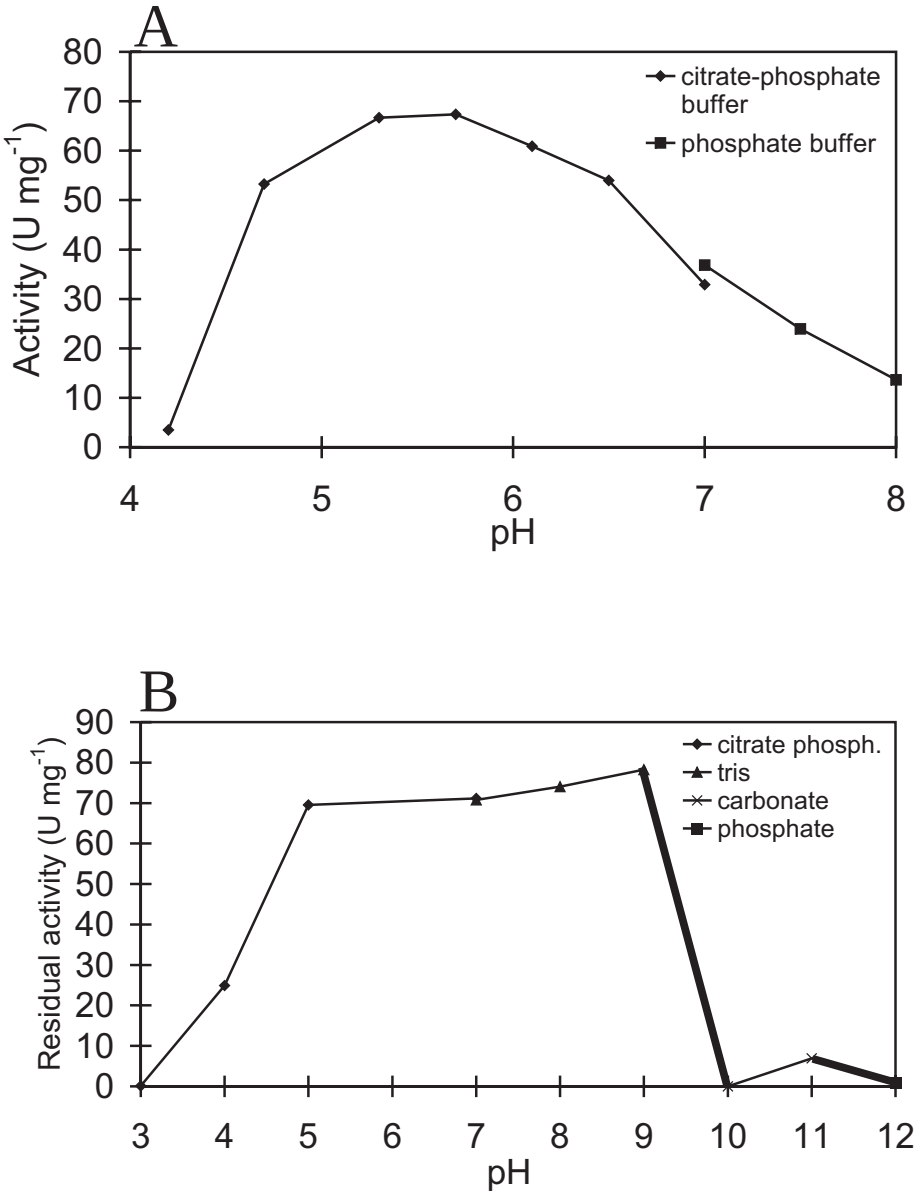


Figure 5.8: pH optimum (A) and stability (B) of CelA. (A) The pH optimum was assayed under standard conditions using 100 mM of the indicated buffers. **(B)** pH stability was determined by incubating the enzyme in the indicated buffers (0.2 M) overnight at 4 C and measuring residual activity under standard conditions.

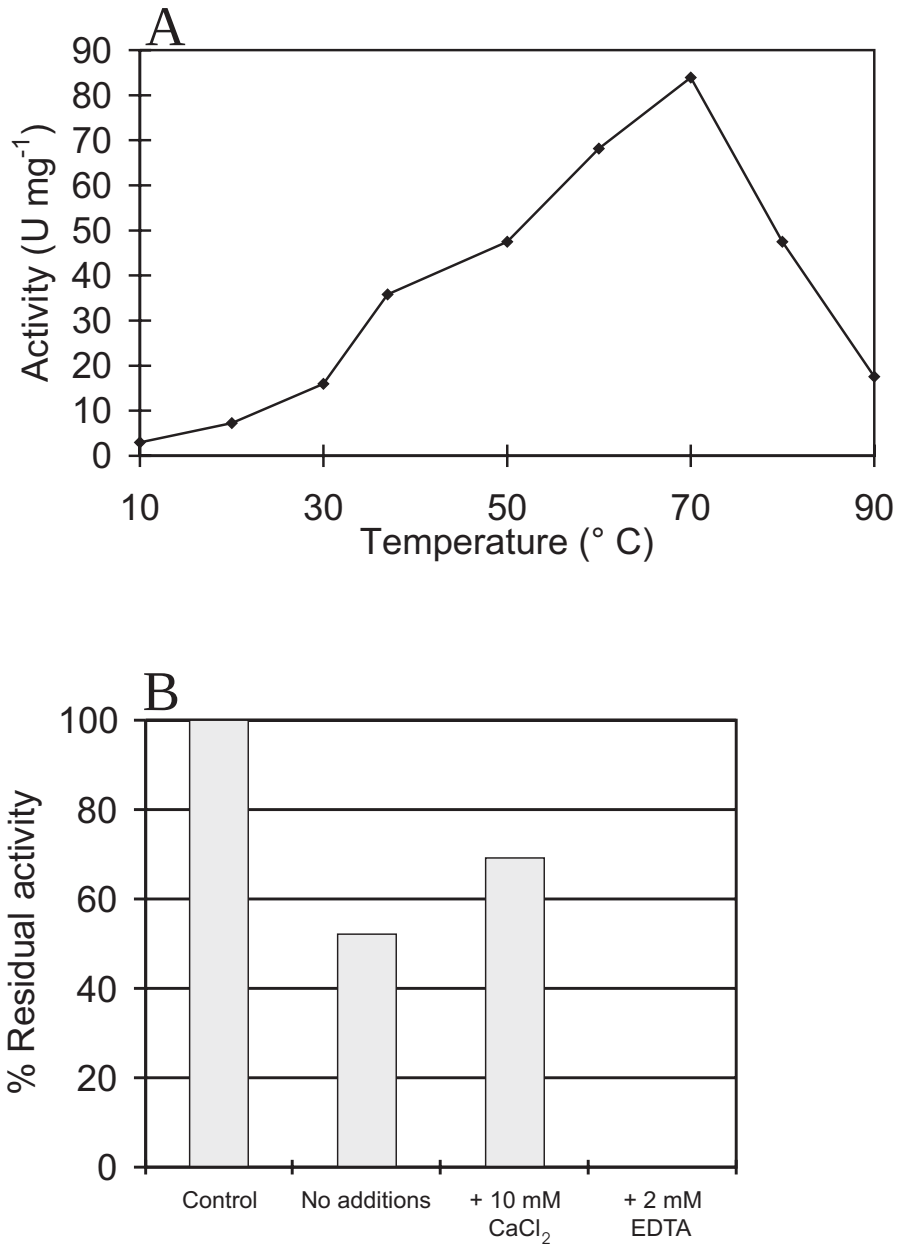


Figure 5.9: Temperature optimum (A) and stability (B) of CelA. (A) The temperature optimum was assayed under standard conditions at the indicated temperatures. **(B)** Temperature stability was determined by incubating concentrated protein at 75 C in buffer with the indicated supplements. After heat treatment, the protein was diluted in buffer containing 10 mM $CaCl_2$. Residual activity was then measured in the dilution buffer under standard conditions, i.e. at 60 C, and compared with the activity of protein which had not been heat treated (control). See materials and methods for details.

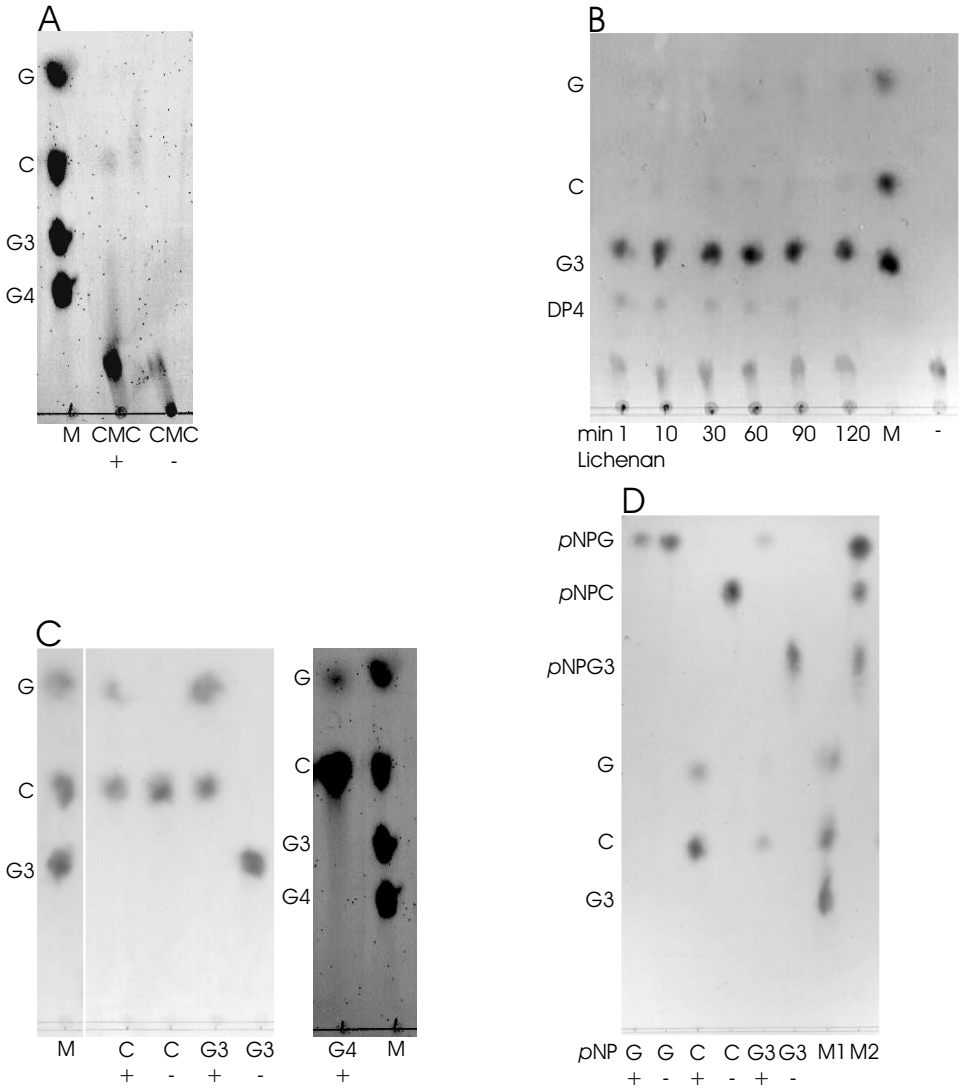


Figure 5.10: TLC analysis of degradation products of various oligosaccharides, lichenan and CMC after incubation with CelA (2.5 **mg ml^{-1}). (A)** Extensive hydrolysis (2 h) of 0.5 % CMC. **(B)** Time course of degradation products of 0.5 % lichenan from 1 to 120 min as indicated. Substrate blank (-) was incubated for 120 min. **(C)** Extensive hydrolysis (2 h) of cellooligosaccharides (10 mM).**(D)** Extensive hydrolysis of aryl-glycosides (10 mM). Substrates were incubated with (+) and without (-) addition of enzyme before TLC analysis. *M*: marker; *G*: glucose; *C*: cellobiose; *G3*: cellotriose; *G4*: cellotetraose; *pNP*: *p*-nitrophenol; *DP*: degree of polymerization;

the initial major product (after 1 minute of digestion) had a DP of three, but minor quantities of smaller and larger (DP4) oligosaccharides were also produced. The latter were considerably diminished after prolonged incubation.

To investigate the activity towards small substrates, the release of pNP from pNP-glycosides was analyzed. The specific activities obtained are shown in tab. 5.2. A striking pattern of hydrolysis emerged, with high chromogenic activity only on pNPC, for which a K_m of 3.0 mM and a k_cat of 7.1 s^{-1} (V_max of 7.13 U mg^{-1}) were determined, resulting in a k_cat / K_m of 2.3 s^{-1} mM^{-1}. Fig. 5.11 shows the corresponding K_m / V_max diagram. Some activity was also detected on pNPG4, but no activity was found on the glucoside and cellotrioside derivatives. To confirm these results, the cleavage products obtained by hydrolysis of chromogenic and unmodified oligosaccharides were identified by TLC (fig. 5.10C, D). The data clearly demonstrate that pNPG was not hydrolysed. Moreover CelA produced mostly cellobiose but also some glucose from the cellotriose analogue pNPC (fig. 5.10D; note that pNP cannot be identified by the TLC system used here). The glucose produced from pNPC hydrolysis must originate from cellobiose breakdown, as no pNPG is produced. Accordingly small amounts of glucose can be produced from cellobiose hydrolysis by CelA (fig. 5.10C). Cellotriose is broken down into similar amounts of cellobiose and glucose (fig. 5.10C). Hydrolysis of pNPG3 produced pNPG, cellobiose and small amounts of glucose (fig. 5.10D). Together with the spectrophotometric measurements which indicate that no pNP is released, this shows that pNPG3 is preferentially hydrolysed by cleaving the cellobiosyl unit at the non-derivatized end. Correspondingly cellotetraose results in formation of mainly cellobiose and little glucose (fig. 5.10C). As no cellotriose is produced, glucose is most likely to result from cellobiose cleavage.

5.1.9 Crystallization of CelA

Crystallization trials carried out with Crystal screens I and II from Hampton research gave only clear wells and precipitate. Initially, drops were made with a protein solution at 3 mg ml^{-1}. To determine which precipitates may be microcrystalline it was tested, which of them dissolve in water. Also precipitates were streak seeded into new solutions (prior to equilibration of the new drop). Seeding with precipitate from Crystal screen II condition 35 (70 % (v/v) 2-methyl-2,4-pentanediol (MPD), 0.1 M Hepes, pH 7.5) resulted in thin needles. Strikingly, neither the needles nor the precipitate dissolved in water or buffer even after two days. Needles were also found in tert-butanol (Crystal screen II condition 40; 25 % (v/v) tert-butanol, 0.1 M Tris pH 8.5), but could not be reproduced. Fig. 5.12, 13, shows the obtained microcrystalline precipitate in tert-butanol. Further trials with other small organic molecules (PEG 400, PEG 600, ethylene glycol) as precipitants yielded only

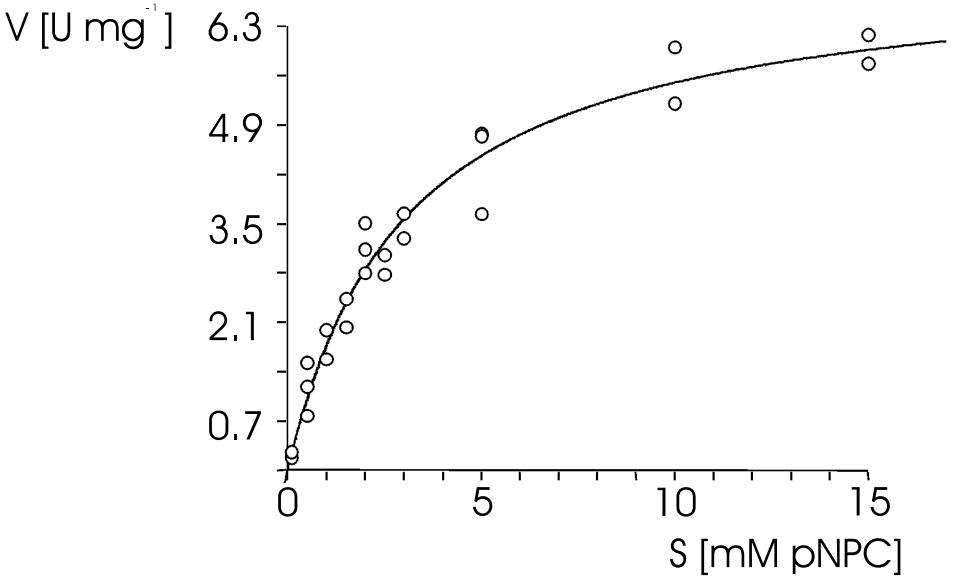


Figure 5.11: K_m / V_{max} diagram of CelA with pNPC as substrate. The curve was fitted with the Marquardt algorithm and a K_m of 3.0 mM and a V_{max} of 7.13 U mg^{-1} were determined.

ethylene glycol (30 % (v/v) ethylene glycol in water, 5 mg ml^{-1} protein) as a promising agent (fig. 5.12, 10–12), giving bundles of needles. However further trials did not improve the quality of these crystals.

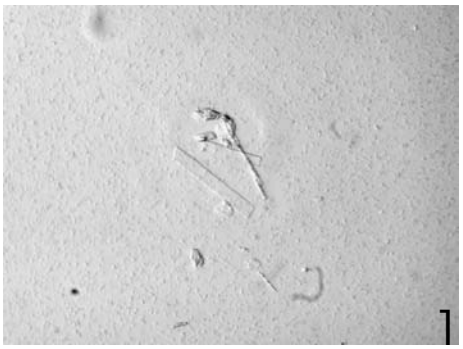
Subsequent efforts concentrated on the crystals obtained in MPD (fig. 5.12, 1–9). Interestingly, trials with the protein dye Izit (Hampton Research) to determine the nature of the crystals, showed that they did not dye even though later diffraction experiments demonstrated them to be protein. However, if the mother liquid was substituted with water the crystals did indeed dye (fig. 5.13).

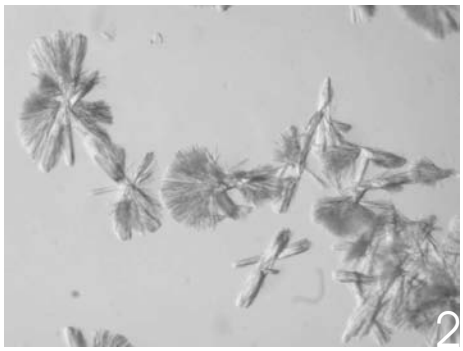
It was found that at a higher initial protein concentration of about 7 mg ml^{-1} there was no need for microseeding. Variation of the crystallization conditions proceeded with testing of different buffers (Tris, Hepes, imidazole) and adjustment of the pH in steps of 0.05. The MPD concentration was also optimized in steps of 1 %. Protein concentration was varied by changing the volumes of protein and MPD in the drop and by varying the protein concentration in the initial protein solution. Finally, kinetics were changed by vapour diffusion rate control (Chayen (1997)). Together these experiments resulted in 57 % MPD, 0.1 M Tris, pH 7.3 and a protein solution with a concentration of 6 mg ml^{-1} as best conditions for growth of CelA crystals. These grew as clusters of needles or thin long plates. They were up to 0.4 mm long and 0.05 mm wide, but were very thin in the third

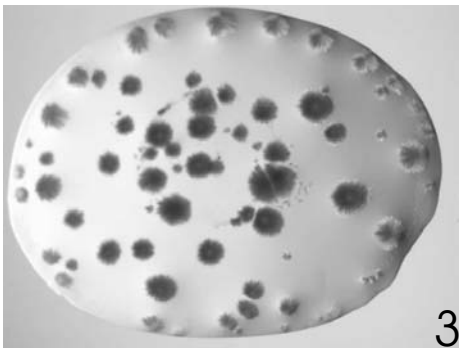
dimension (fig. 5.12, 5). At higher MPD or protein concentrations an oil was produced. Tests later carried out with the crystals (Heidi A. Ernst and Leila Lo Leggio, Eckert et al. (2003)) showed that they diffracted weakly and were disordered in one direction. After testing several crystals, one was found which gave ordered diffraction in all directions. A full data set was collected to 3.0 Å resolution, but because of the limited homology of CelA and CelD from *C. thermocellum*, for which the 3D structure is known, the structure could not be solved by molecular replacement.

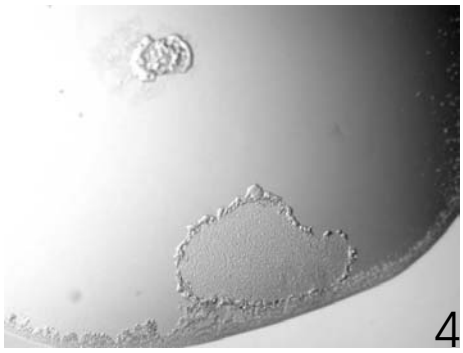
Employment of the Hampton Research additive screens 1, 2 and 3 showed that the reducing agent 1,4-dithio-DL-threitol (DTT) made the needles obtained in MPD grow faster (over night instead of two or three days). L-cysteine, which also contains a thiol group and spermine tetra-HCl, a polyamine, gave crystals of a different morphology. Addition of 0.01 M cysteine giving rise to small bunches of needles and spherolites (fig. 5.12, 14) and 0.01 M spermine showing twisted bunches of small needles (fig. 5.12, 15). Addition of cellobiose resulted in inferior crystals at a concentration of 10 mM and produced an oil at higher concentrations (fig. 5.12, 4). Ruthenium Red (0.5 mM) led to precipitate (fig. 5.12, 9).

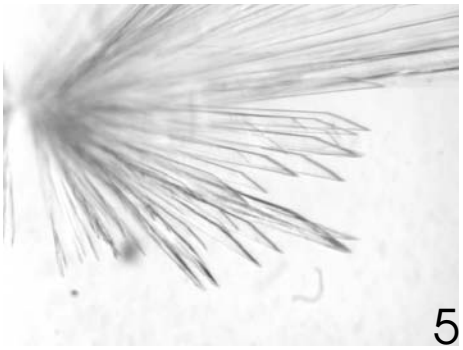
Figure 5.12: Results of hanging-drop vapour-diffusion crystallization trials with CelA under standard conditions.
No., CelA conc., precipitant, buffer, additive and remarks are given:
1, 5 mg ml^{-1}, 40 % MPD, 0.1 M Tris, pH 7.3, seeding;
2, 5 mg ml^{-1}, 49 % MPD, 0.1 M Tris, pH 7.3, seeding;
3, 16 mg ml^{-1}, 50 % MPD, 0.1 M Tris, pH 7.3;
4, 6 mg ml^{-1}, 54 % MPD, 0.1 M Tris, pH 7.3, 0.1 M cellobiose;
5, 6 mg ml^{-1}, 57 % MPD, 0.1 M Tris, pH 7.3;
6, 5 mg ml^{-1}, 57 % MPD, 0.1 M Tris, pH 7.3, seeding;
7, 5 mg ml^{-1}, 60 % MPD, 0.1 M Tris, pH 7.3;
8, 16 mg ml^{-1}, 60 % MPD, 0.1 M Tris, pH 7.3;
9, 5 mg ml^{-1}, 60 % MPD, 0.1 M Tris, pH 7.3, 0.5 mM Ruthenium red;
10, 5 mg ml^{-1}, 30 % EG, water, seeding;
11, 5 mg ml^{-1}, 30 % EG, water, 10 mM cellobiose, seeding;
12, 5 mg ml^{-1}, 35 % EG, water;
13, 5 mg ml^{-1}, 20 % tert-butanol, 0.1 M Na citrate, pH 5.6;
14, 5 mg ml^{-1}, 55 % MPD, 0.1 M Hepes, pH 7.5, 0.01 M L-cysteine;
15, 5 mg ml^{-1}, 55 % MPD, 0.1 M Hepes, pH 7.5, 0.01 M spermine tetra-HCl.
MPD, 2-methyl-2,4-pentanediol; EG, ethylene glycol.

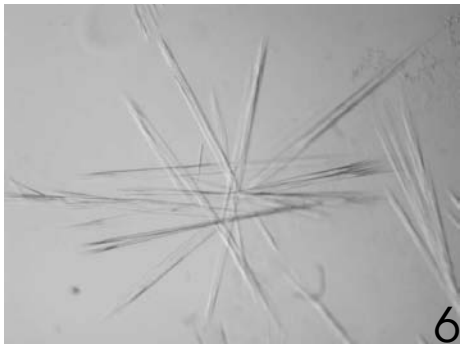

1

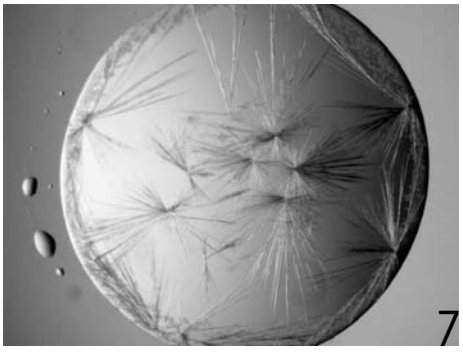

2

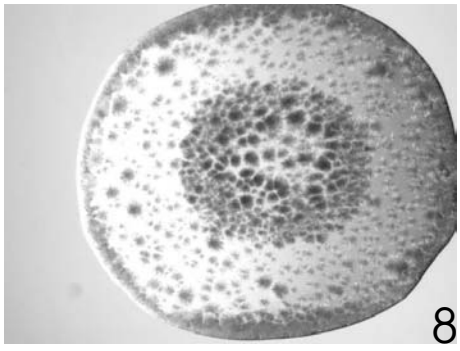

3


4


5


6


7


8

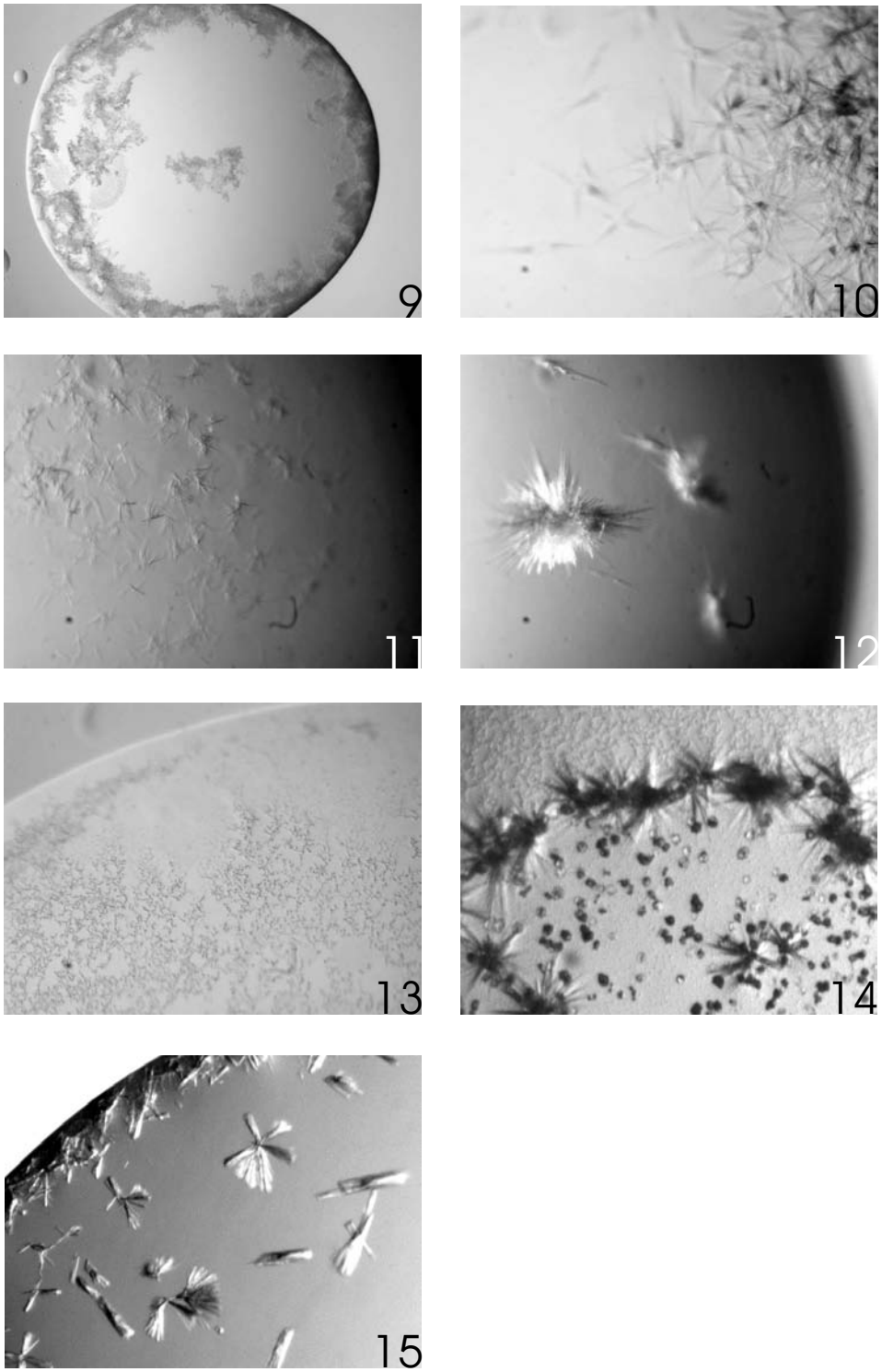


Figure 5.12: Continued

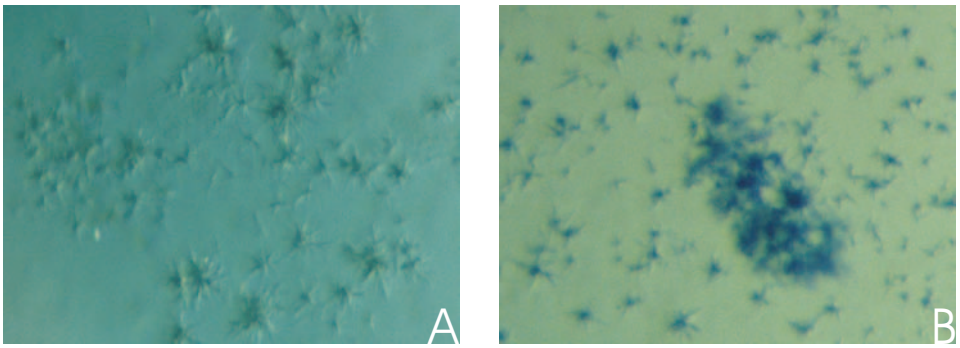


Figure 5.13: Dying of CelA crystals with Izit. (A) Crystals did not dye in the mother liquor. However if the solution was replaced with water and Izit added, the crystals dyed (B).

5.2 Gene cloning and characterization of the cell-associated enzyme CelB

5.2.1 Purification and partial sequencing of wild-type CelB

In the initial stage of this work, *A. acidocaldarius* was screened for extracellular, thermoacidophilic enzymes with polysaccharide-degrading activities. The organism was found to utilize a variety of sugars including xylan as sole source of carbon and energy. However, no xylanase activity was detected in the culture supernatant. Therefore, a cell-associated enzyme with the catalytic domain facing the extracellular medium seemed plausable. Cells were treated with the detergent Triton X-100 to remove cell-bound proteins and the extract (sec. 4.3.3) was analyzed with regard to xylanase activity. Indeed, xylan-degrading activity was detected. The xylanase activity remained cell-bound even after the culture reached the stationary phase of growth. SDS-PAGE of Triton-extracted proteins followed by silver staining revealed about ten major proteins with molecular masses ranging from 30 to 100 kDa (fig. 5.14A, lane 1). Zymogram analysis demonstrated that at least three of these proteins displayed activity towards oat spelt xylan and CMC (fig. 5.14A, lanes 2 and 3).

Further efforts were concentrated on the 100 kDa protein. Purification of the wild-type protein proved to be difficult, because it was produced only in low amounts by the bacterium and did not elute from ion exchange columns as a single peak, but was spread out over a large volume of eluant. This was put down to non-specific interaction of the protein with the column material due to the hydrophobic nature of the protein. However, purification eventually was achieved by ion exchange chromatography using Q-Sepharose in the presence of 0.05 % Triton X-100. Elution took place with a large volume to minimize loss and pure fractions were selected after analysis on silver-stained SDS gels. Fig. 5.15 shows samples from such a chromatography. From a 1 litre culture of *A. acidocaldarius* about 2.0 mg of the 100 kDa protein, exhibiting on average a CMCase activity of 10.3 U mg^{-1} and a xylanase activity of 0.9 U mg^{-1}, were routinely obtained (fig. 5.16 A, lane 1).

N-terminal sequence analysis of the protein revealed the peptide sequence:

```
        10  14
ADVXSTPIXmexqv
```

'X' denotes uncertain positions, while small letters apply to amino acids giving a weakened signal during sequencing.

Analysis of a peptide fragment generated by cyanogen bromide gave rise to the sequence:

```
          12
MVAELXREINAY
```

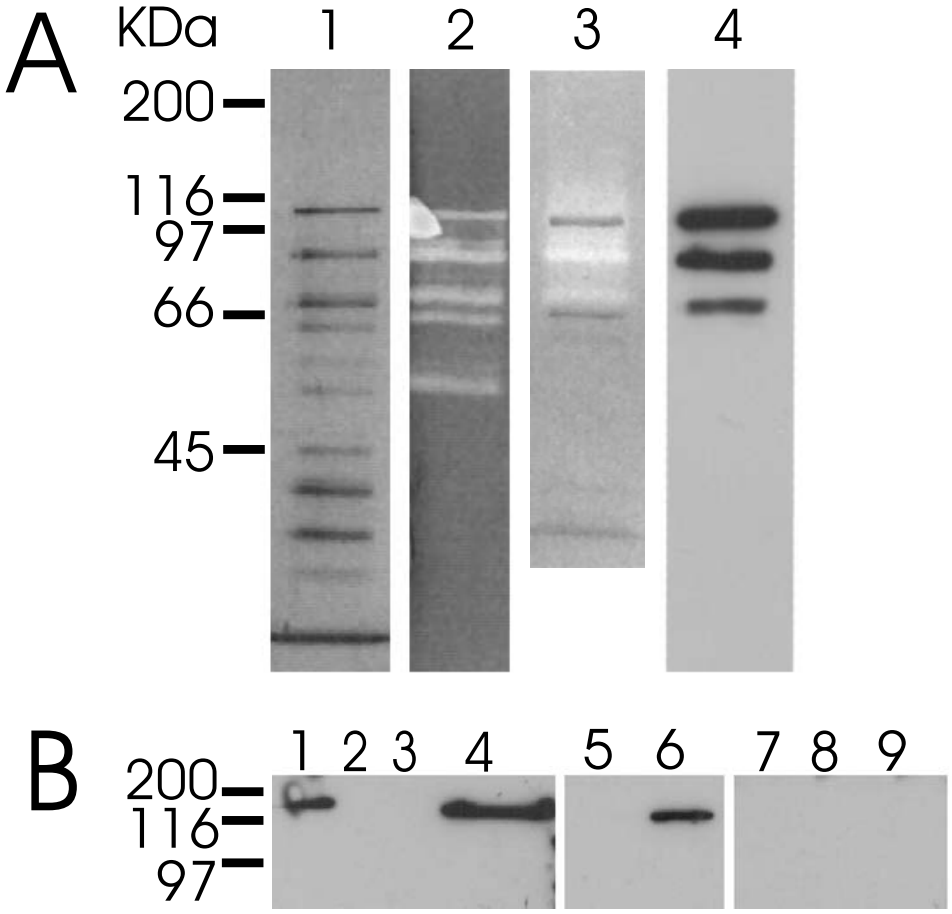


Figure 5.14: Identification of CelB in Triton extract from *A. acidocaldarius*. **(A)** Triton extract (25 l per lane) from cells grown on oat spelt xylan after SDS-PAGE. *Lane 1*, silver staining; *2*, zymogram analysis with CMC; *3*, silver staining and zymogram analysis with oat spelt xylan; *4*, Western blotting with antibodies raised against wild-type CelB. **(B)** Western blots of Triton extracts (25 l) from *A. acidocaldarius* grown on different substrates. *Lane 1*, cellobiose; *2*, starch; *3*, arabinan; *4*, birchwood xylan; *5*, xylose; *6*, CMC; *7*, glycerin; *8*, glucose; *9* maltose.

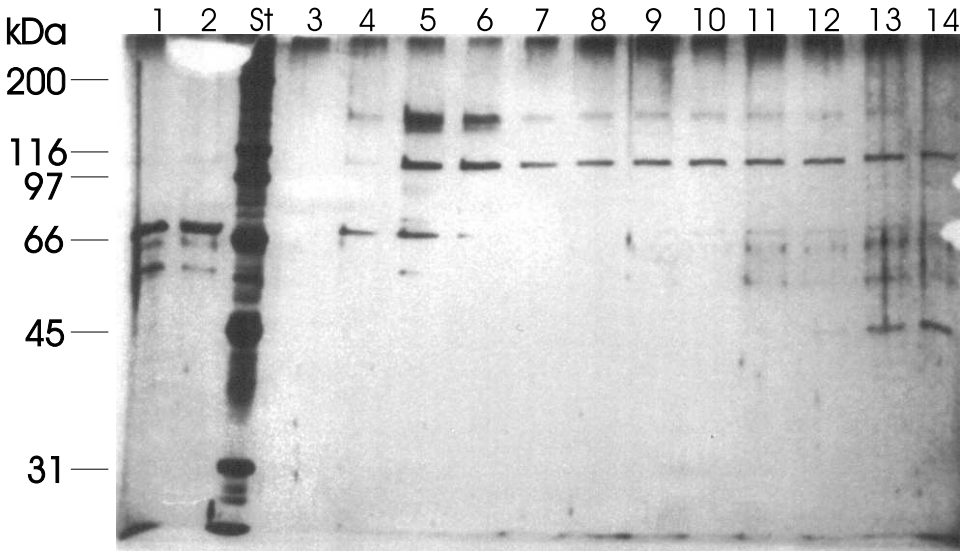


Figure 5.15: Purification of wild-type CelB by Q-sepherose column chromatography. 15 l of each fraction were subjected to SDS-PAGE and silver staining. *St*, protein size standard; *lane 1*, Triton extract from *A. acidocaldarius* (450 ml total volume); *2*, pooled flow-through and wash fraction (total volume 600 ml); *3–14*, salt gradient (0 to 0.2 M NaCl) fractions as CelB elutes (500 ml total volume, 10 ml per fraction). Fractions corresponding to lanes 7–12 were used for further purification. See sec. 4.3.3 and sec. 5.2.1 for details.

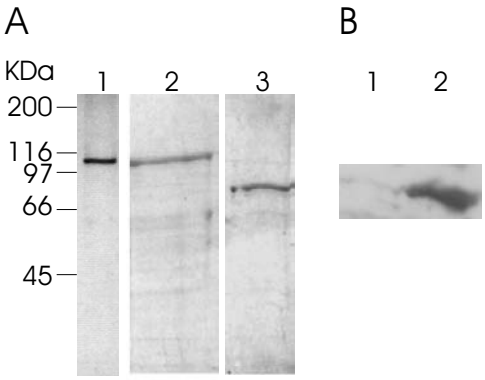


Figure 5.16: (A) SDS-PAGE of purified wild-type and recombinant forms of CelB. *Lane 1*, wild-type CelB (0.2 g), silver stained; *2*, full-length recombinant CelB (3 g), stained with Serva Blue R; *3*, recombinant $\text{CelB}_{\text{trunc}}$ (3 g), stained with Serva Blue R. **(B) Reaction of antibodies raised against WT CelB with recombinant $\text{CelB}_{\text{trunc}}$.** Western blot of whole-cell extracts (25 l) of *E. coli* TOP10 (*1*, control) and *E. coli* TOP10 (pKE2201) (*2*) after expression of $celB_{trunc}$.

whereby methionine at the beginning of the fragment is assumed due to the cleavage mechanism of cyanogen bromide.

No homolgy to an entry in the data base was found using BLASTP.

5.2.2 Expression of *celB* in *A. acidocaldarius*

The purified 100 kDa protein was injected into rabbits to raise polyclonal antibodies. Subsequent immuno blot analysis of the Triton extract revealed that, in addition to the 100 kDa protein, two other protein bands with a molecular mass of $\sim$ 70 and 80 kDa strongly crossreacted with the antiserum (fig. 5.14A, lane 4). This result may imply that these proteins represent degradation products of the 100 kDa protein. However, further bands in the zymogram are likely to be the result of additional proteins that participate in the complete degradation of CMC or xylan. Furthermore, Western blot analysis of Triton extracts from *A. acidocaldarius* cells grown on different substrates also showed that, in addition to oat spelt xylan, synthesis of the 100 kDa protein was induced by birchwood xylan, CMC, and cellobiose, but not by glycerol, glucose, xylose, maltose, starch or arabinan (fig. 5.14B).

5.2.3 Cloning and sequence analysis of the *celB* gene

Because previous efforts had not led to any of the genes coding for enzymes of the Triton extract a further attempt was undertaken to clone xylan or cellulose degrading enzymes. And since zymogram analysis had shown that the major enzymes were active against CMC as well as xylan and CMC gave clearer lysis zones, this substrate was used to screen for activity. In addition, a high concentration of buffer (250 mM) was used in the top agar, because the screening of the plasmid gene bank had yielded a gene coding for a neutrophilic enzyme (CelA) even though a buffer at pH 3.5 had been present in the top agar.

Instead of a plasmid gene bank a λ derived vector was chosen. The advantages being a larger insert size and cell lysis by the phage. Larger inserts result in fewer number of clones necessary for a representative gene bank and a higher chance of a full-length clone. Cell lysis by the phage leads to exposure of heterologous proteins to the substrate in the top agar even if the signal peptide is not recognized by the host.

The cloning procedure with the Zap Express vector yielded a gene bank with $2.3 \cdot 10^6$ pfu in total of which $3 \cdot 10^4$ were background ('blue') plaques. Insert sizes ranged from $\sim$ 3–10 kb. Screening of 45,000 plaques for CMC activity with the substrate overlay method and subsequent excision resulted in 5 phagemids. One clone harboured the previously described intracellular cellulase CelA as identified by Western blotting, but a second clone, containing phagemid pKE25, reacted with antibodies raised against the wild-type 100 kDa protein. Nucleotide sequencing revealed an incomplete ORF

which coded for a protein that displayed high sequence similarities with endoglucanases and arabinofuranosidases. Eventually, screening of digested chromosomal DNA by Southern hybridization with a fragment from the 3' end of the incomplete ORF gave rise to an overlapping clone, harbouring plasmid pKE25a5, that contained the rest of the ORF (fig. 5.17A).

The complete ORF encoded a preprotein of 959 amino acids with a molecular mass of 100.849 kDa. A possible start codon (TTG) with a putative ribosome-binding site was found together with possible −10 (TATAAC) and −35 (TTGACA) regions. A probable catabolite responsive element (CRE) and an inverted repeat were found between the promotor and the ORF. The CRE sequence showed only one mismatch in the last position when compared to the consensus sequence TGT/AAANCGNTNA/TCA (Hueck and Hillen (1995)). The inverted repeat may generate a 10 bp stem with only one mismatch and a loop of 4 nt (fig. 5.17 B). SignalP (Nielsen et al. (1997)) detected a possible signal peptide whose cleavage site was located C-terminal of amino acid 25 of the preprotein (fig. 5.17 B). 19 amino acids C-terminal of the cleavage site a sequence was found which displayed 79 % identity with the sequence obtained from the N-terminus of the purified wild-type 100 kDa protein (fig. 5.17 B):

```
N-term. 100 kDa protein: 1 ADVXSTPIXMEXQV 14
CelB:                   43 ADVVSTPISMEIQV 56
```

Moreover, residues 485–496 of the translated ORF showed only one substitution when compared with the internal sequence of the 100 kDa protein:

```
CNBr fragment, 100 kDa protein: 1 MVAELXREINAY 12
CelB:                         485 MVAELKREINAY 496
```

In both cases, amino acids which differed were at positions were amino acid sequencing information was uncertain. Taken together, we conclude that the ORF indeed codes for the 100 kDa protein that was purified from *A. acidocaldarius* cells. The ORF was designated *celB*.

The *celB* gene is flanked by two divergently transcribed putative ORFs, encoding a LacI / GalR type transcription regulator (152 nucleotides upstream of *celB*) and a LysE type exporter (176 nucleotides downstream of *celB*, fig. 5.17A). Further downstream putative ORFs with homologies to a sensory transduction histidine kinase and an ABC transporter were detected. Upstream of the *lacI* gene a putative permease is flanked by two transcriptional regulatory proteins. Details are given in tab. 5.3 and fig. 5.18.

A database search using BLASTP revealed highest sequence similarity of CelB (28 % identity, 45 % homology over a length of 410 amino acids)

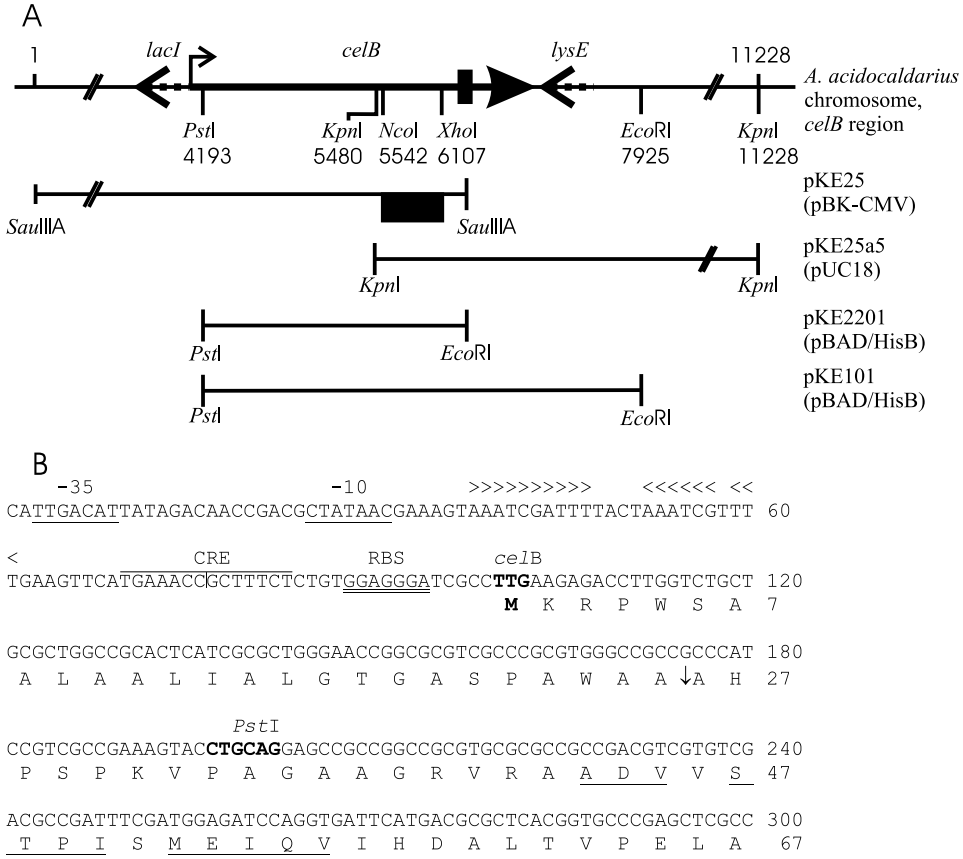


Figure 5.17: (A) Overview of the *celB* region and cloning strategy. Shown is the *celB* region of the *A. acidocaldarius* chromosome (*top line*). *Numbers* indicate nucleotide positions relative to the 5' *Sau*IIIA site of the original clone (pKE25). ORFs are represented by *arrows* in the direction of transcription. *Dashed arrows* show ORFs neighbouring *celB* with putative assignments. The *crooked arrow* indicates the *celB* promotor detailed in B. The *thick vertical bar* indicates the end of the ORF in $celB_{trunc}$. Restriction sites relevant to the cloning strategy are given. At the bottom inserts of the constructed plasmids are drawn in relation to the *celB* region with the restriction sites used for excision of the insert prior to ligation in the host vector (in *parentheses*). The DNA fragment of pKE25 used for Southern hybridization is marked by a *black box*. **(B) Sequence analysis of the 5' region of the *celB* gene.** Shown are the nucleotide sequence and the corresponding amino acids. Indicated for the nucleotide sequence are the putative −10 and −35 promotor regions (*underlined*), an inverted repeat (>< above paired bases), a putative CRE (*overlined* with the axis of symmetry denoted by a *vertical line*), the ribosome-binding site (*double-underlined*), the start codon (*boldface*) and the *Pst*I site used for subcloning (*boldface*). Indicated in the amino acid sequence are the putative cleavage site of the signal peptidase (*arrow*) and the amino acid sequence found in the N-terminus of the wild-type protein (identical positions *underlined*).

Alicyclobacillus acidocaldarius, celB contig Putative function	Nt pos.	No. of aa	Organism / gene	Homology
transcriptional repressor / activator	146–565	139	*Bacillus sp.* RC607, *merR*	38 %
permease	800–2338	512	*Sulfolobus solfataricus*, NP_342797	72 %
transcriptional antiterminator	2359–2910	183	*Bacillus subtilis*, *glpP*	70 %
LacI / GalR type transcriptional regulator	2907–3920	337	*Escherichia coli*, *purR*	48 %
CelB	4096–6975	959	see sec. 5.2.3 for details	
LysE type exporter	7128–7787	219	*Salmonella typhimurium* LT2, *rhtC*	26 %
sensory transduction histidine kinase	∼ 8048–9650	∼ 530	∼ 70 % / 265 aa	
ABC-transporter	∼ 10064–10646	∼ 580	∼ 70 % / 100 aa	

Table 5.3: Contig containing *celB* with putative assignments of ORFs. Nt positions (*Nt pos.*) of the ORFs in the contig are given in relation to the 5' *Sau*IIIA site of the original clone (pKE25, fig. 5.17A) and the no. of amino acids of the respective proteins are indicated (*No. of aa*). The *putative function* of the assumed proteins and their amino acid sequence similarities (*Homology*) with proteins found in the database using BLASTP are given. Organism, GenBank no. or gene names of these proteins are indicated (*Organism / gene*). Where sequence data was not of high quality only approximate positions and similarities with various proteins are given (∼). See also fig. 5.18.

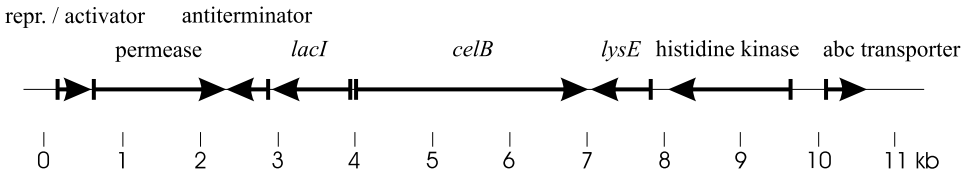


Figure 5.18: Overview of the *celB* containing contig. *Numbers* indicate nucleotide positions in kb relative to the 5' *Sau*IIIA site of the original clone (pKE25). ORFs are represented by *arrows* in the direction of transcription with putative assignments. See also tab. 5.3.

with endoglucanase F (EGF) from *Fibrobacter succinogenes* S85 (Malburg et al. (1997), GenBank acc. no. U39070) which belongs to family 51 of glycoside hydrolases. Among the 32 other matches found 19 were arabinofuranosidases. After three iterations PSI-BLAST showed only three proteins not classified as arabinofuranosidases among the top 30 matches. Sequence comparison of CelB with all members of GH51 revealed a central catalytic domain ranging from amino acids Thr 223–Pro 702. Fig. 5.19 shows an alignment of this domain with the catalytic domain of EGF. Catalytic residues have been inferred from sequence alignments in this family (Zverlov et al. (1998)) and experimentally confirmed (Beylot et al. (2001)). The conserved motif GNE is also present in CelB identifying Glu 366 as the acid / base catalyst. However, the catalytic nucleophile could not be identified unequivocally from the sequence. Glu 510 is a possible candidate (fig. 5.19). A phylogenetic tree constructed from the alignment of the catalytic domains showed that CelB forms a distinct cluster with EGF (fig. 5.20). These two are the only enzymes characterized as endoglucanases in GH family 51.

5.2.4 Purification of recombinant CelB and CelB$_{\mathbf{trunc}}$

It was possible to purify recombinant full-length CelB by Ni-NTA affinity chromatography provided that 0.1 % Tween 20 or 0.5 % Triton X-100 were present throughout the purification procedure to keep the protein in solution. Routinely, 18 mg of purified CelB exhibiting a CMCase activity of 11.4 U mg^{-1} (average of two independent preparations) were obtained from a 1 litre culture of *E. coli* TOP10 (pKE101) (fig. 5.16 A, lane 2).

Full length recombinant CelB could only be purified with a low yield. As was the case with the wild-type protein elution from the column was hampered, probably due to hydrophobic interactions of the protein with the column material. Also, high concentrations of detergents led to precipitation in the reducing sugar assay, making it necessary to dilute the samples prior to absorbance measurements. These difficulties made a soluble pro-

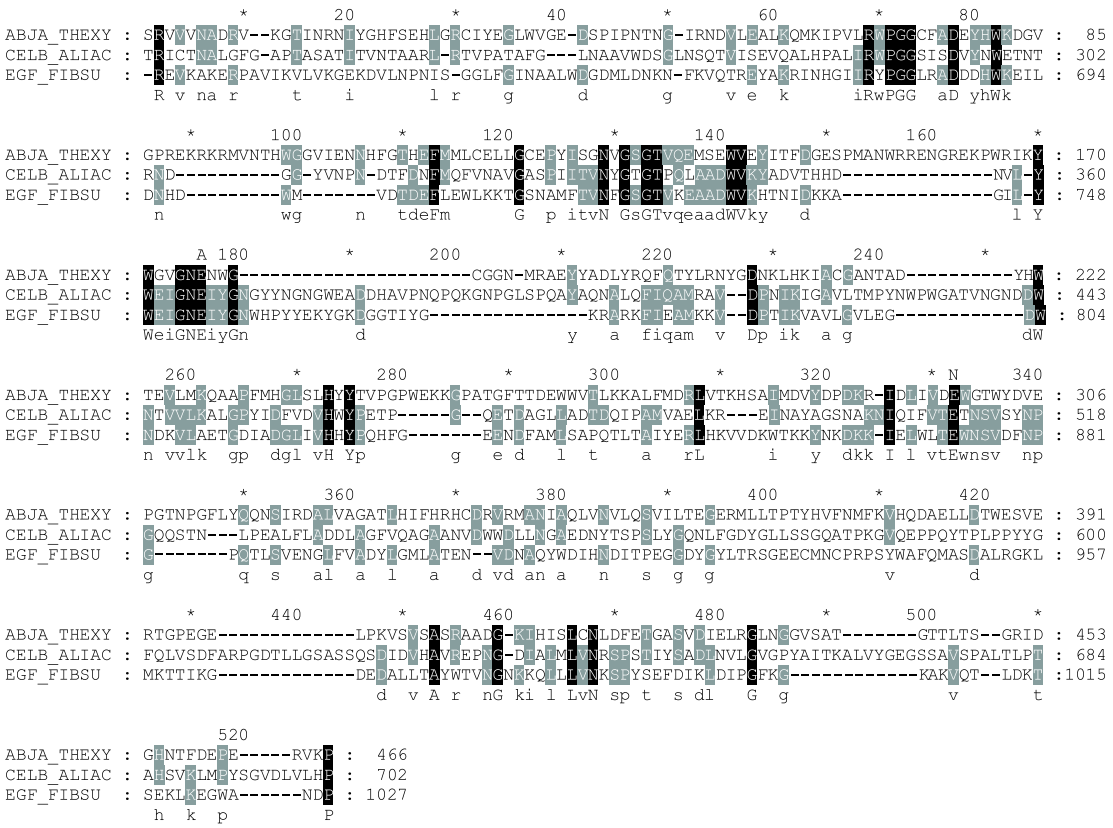


Figure 5.19: Sequence alignment of the proposed catalytic domain of CelB (CELB_ALIAC) and those of two other representatives of family 51 glycoside hydrolases, AbjA from *Thermobacillus xylanilyticus* (ABJA_THEXY) and EGF from *Fibrobacter succinogenes* (EGF_FIBSU). The alignment was carried out using the program ClustalX and edited by eye. The sequences continue beyond the shown alignment. Identical residues in two and all three proteins are high-lighted in *gray* and *black*, respectively. Catalytic acid / base and catalytic nucleophile are marked by *A* and *N*, respectively.

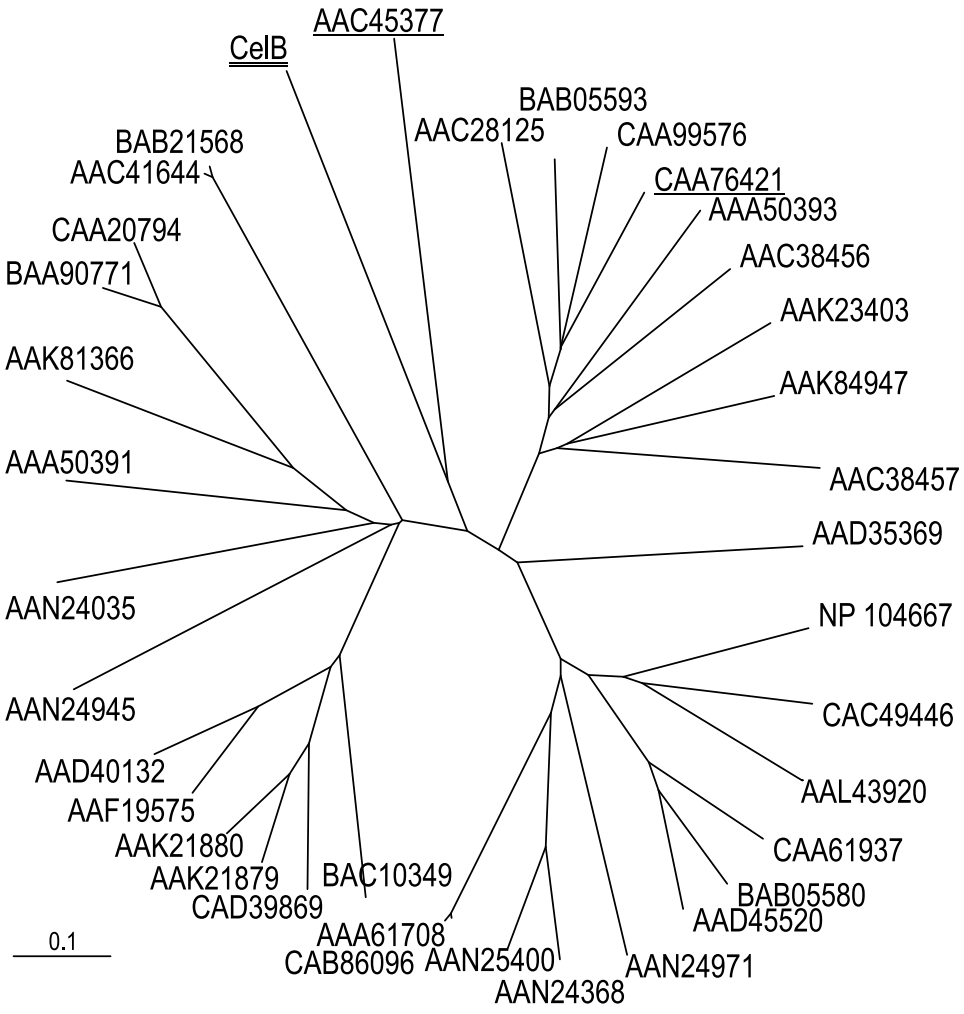


Figure 5.20: Phylogenetic tree of catalytic domains belonging to GH family 51. CelB (*doubly-underlined*) forms a distinct group with EGF (*underlined*, AAC45377). Also *underlined* is AbjA (CAA76421) from the thermophile *T. xylanilyticus*. Bar length, extent of exchange of 0.1 per residue. GenBank / GenPept accession codes are given. *Agrobacterium tumefaciens* C58: AAL43920, ORF Atu3104; *Arabidopsis thaliana*: AAD40132, ORF At5g26120/T1N24.13; AAF19575, ORF At3g10740/T7M13_18; *Aspergillus niger*: AAC41644, arabinofuranosidase A; *A. niger var. awamorii*: IFO4033,BAB21568, ArfA; *Bacillus halodurans* C-125: BAB05580, ORF AbfA (BH1861); BAB05593, ORF Xsa (BH1874); *B. subtilis subsp. subtilis* str. 168: CAA61937, arabinofuranosidase 1; CAA99576, arabinofuranosidase 2; *Bacteroides ovatus*: AAA50391, arabinosidase 1; AAA50393, arabinosidase 2; *Bifidobacterium longum* NCC2705: AAN24035, BL0181; AAN24368, AbfA1; AAN24945, BL1138; AAN24971, AbfA2; AAN25400, AbfA3; *Caulobacter crescentus* CB15: AAK23403, ORF CC1422; *Cellvibrio japonicus*: AAK84947, arabinofuranosidase; *Clostridium acetobutylicum* ATCC824: AAK81366, ORF CAC3436; *C. stercorarium*: AAC28125, arabinofuranosidase; *Cytophaga xylanolytica*: AAC38456, arabinofuranosidase I; AAC38457, arabinofuranosidase II; *Geobacillus stearothermophilus* T-6: AAD45520, abfA; *Hordeum vulgare*: AAK21879, AXAH-I; AAK21880, AXAH-II; *Mesorhizobium loti* MAFF303099: NP 104667, Mll3591; *Oryza sativa*: BAC10349, OJ1200$_C$08.20; CAD39869, OSJNBb0058J09.6; *Sinorhizobium meliloti* 1021: CAC49446, AbfA; *Streptomyces chartreusis*: BAA90771, arabinofuranosidase I; *S. coelicolor* A3(2): CAA20794, ORF SCI35.05c; CAB86096, AbfA; *S. lividans* 66: AAA61708, AbfA; *Thermotoga maritima*: AAD35369, ORF TM0281.

tein desirable. Because a hydropathy plot of CelB (fig. 6.3) indicated that the hydrophobic nature of the protein was due to the C-terminal part of the protein, the truncated protein $CelB_{trunc}$ was expressed and purified. Indeed, $CelB_{trunc}$, lacking the C-terminal 203 amino acids was soluble in the absence of detergent and could therefore be purified from *E. coli* strain TOP10 (pKE2201) by essentially the same procedure but without Triton X-100 or Tween (fig. 5.16 A, lane 3). The protein displayed a similar activity towards CMC of 10.1 U mg^{-1} (average of two independent preparations), strongly suggesting that the C-terminal portion is dispensable for activity. Thus, further characterization of the enzymatic properties was carried out predominantly with $CelB_{trunc}$.

Heterologous CelB and $CelB_{trunc}$ were bound by antibodies directed against the wild-type 100 kDa protein, giving further evidence to the hypothesis, that *celB* encodes the 100 kDa protein (fig. 5.16 B).

5.2.5 pH and temperature dependency of CelB

Wild-type CelB showed a pH optimum of 4 in accordance with the acidic surroundings of *A. acidocaldarius*. At pH 3 and 5 it still displayed 50 % activity, while no activity was recorded at pH values below 2 or above 6. $CelB_{trunc}$ behaved similarly (fig. 5.21A) and also displayed remarkable tolerance to acidic pH, being stable (80 % residual activity) overnight at pH values ranging from 1.5 to 7. Increasing alkalinity above pH 11 completely and irreversibly inactivated the enzyme (fig. 5.21B).

Adaptation to high temperatures could be observed for the wild-type protein and $CelB_{trunc}$, both of them displaying a temperature optimum of 80 C. However, the optimum curve of the truncated protein was broader than that of the wild-type for unknown reasons. 50 % activity was found at 90 C. No activity was found at 100 C or below 40 C (fig. 5.22A). When temperature stability of $CelB_{trunc}$ was measured, a 10 minute incubation at 90 C prior to determination of residual activity resulted in complete inactivation of the enzyme in contrast to the activity found at the same temperature. After incubation for 1 h at 80 C residual activity was still 60 % of the control (activity measured without heat treatment) and, remarkably, treatment at 70 C even resulted in a higher activity (fig. 5.22B).

5.2.6 Enzymatic properties and substrate specificity

Determination of kinetic parameters of $CelB_{trunc}$ using CMC as substrate yielded a K_m of 0.35 mg ml^{-1} and a V_{max} of 10.8 U mg^{-1}, resulting in a k_{cat} of 881 min^{-1} (fig. 5.23).

Under standard conditions, optimal activity was obtained in the presence of Ca^{2+} and Mg^{2+} ions (10 mM each). Omitting the bivalent cations decreased the activity by 50 %. A small stimulating effect in the presence

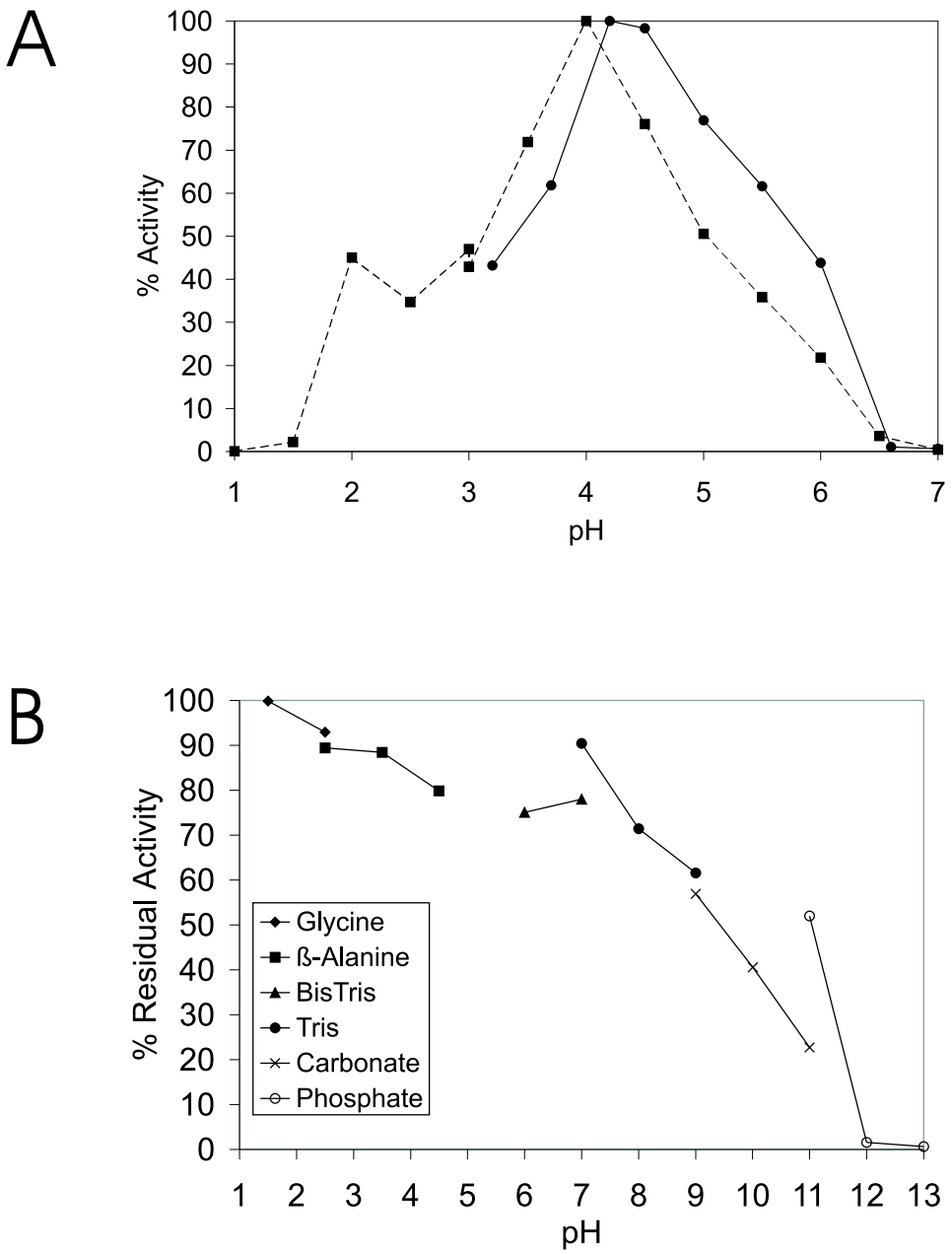


Figure 5.21: (A) pH optima of wild-type CelB (*circles*) and recombinant $\mathbf{CelB_{trunc}}$(*squares*). Cellulase activity was assayed under standard conditions at 70 C at the indicated pH values. Glycine (pH 1–3) and citrate-phosphate (pH 3–7) were used as buffers. Control activities (100 %) for wild-type CelB and $CelB_{trunc}$ were 12.2 U mg^{-1} and 10.8 U mg^{-1}, respectively. **(B) pH stability of $\mathbf{CelB_{trunc}}$.** Enzyme was incubated overnight at 4 C in the indicated buffers after which residual activity was assayed under standard conditions. Control activity (100 %) was 8.7 U mg^{-1}.

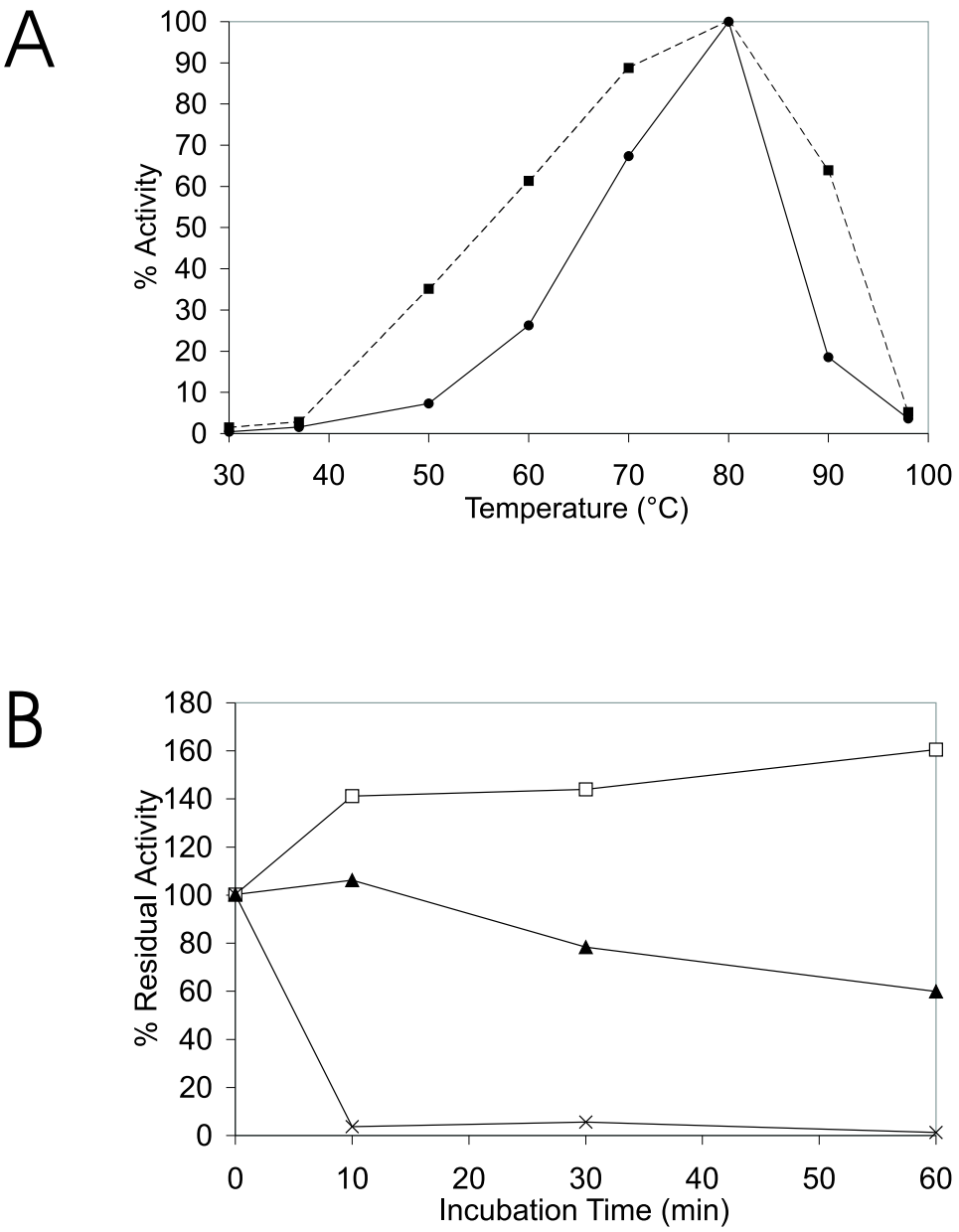


Figure 5.22: (A) Temperature optimum of wild-type CelB (*circles*) and recombinant $\mathbf{CelB_{trunc}}$(*squares*). Cellulase activity was assayed at pH 3.5 for 30 min. Control activities (100 %) for wild-type CelB and $CelB_{trunc}$ were 8.8 U mg^{-1} and 8.1 U mg^{-1}, respectively. **(B) Temperature stability of $\mathbf{CelB_{trunc}}$.** Enzyme was incubated at 70 (*squares*), 80 (*triangles*) or 90 C (*crosses*) for the indicated times and residual activity was measured under standard conditions. Control activity prior to heat treatment (100 %) was 11.7 U mg^{-1}.

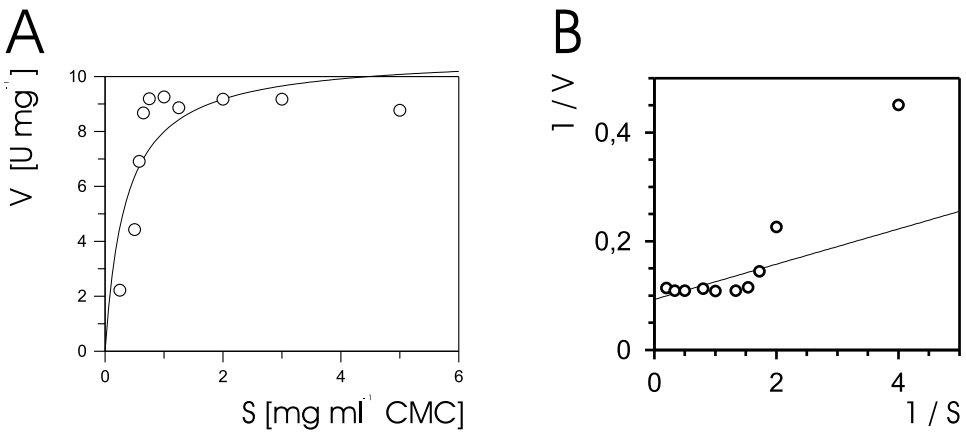


Figure 5.23: (A) K_m / V_{max} diagram of CelB with CMC as substrate. The curve was fitted with the program Grafit and a K_m of 0.35 mg ml^{-1} and a V_{max} of 10.8 U mg^{-1} were determined. **(B) Lineweaver-Burk plot of the same data.**

of these cations was also described for AbfA from GH family 51 (Gilead and Shoham (1995)). In contrast, 10 mM Zn^{2+} caused a 78 % inhibition of $CelB_{trunc}$ activity. Inhibition by Zn^{2+} is typical of many GHs (Blanco et al. (1995), Dsterhft et al. (1997)), but not of all members of family 51 (Debeche et al. (2000), Gilead and Shoham (1995)).

Apart from CMC, $CelB_{trunc}$ was found to hydrolyze phosphoric acid-swollen cellulose (0.81 U mg^{-1}) and, like the wild-type protein, showed activity towards oat spelt xylan (0.6 U mg^{-1}). In contrast, no activity was found with crystalline cellulose (Avicel PH101), birchwood xylan, starch and, most strikingly, with arabinan, linear arabinan or pNPAraf, in spite of the described sequence similarity to other arabinofuranosidases. In order to discriminate an endo or exo mode of action degradation of CMC and oat spelt xylan by $CelB_{trunc}$ was analysed by TLC. A time course showed that in the initial reaction only high molecular mass products were released from oat spelt xylan (fig. 5.24A). The appearance of disaccharides (xylobiose in the case of oat spelt xylan and cellobiose in the case of CMC) as final degradation products was only observed at a high enzyme concentration (80 g ml^{-1}) (fig. 5.24B, C). To confirm the endowise action of the enzyme, its hydrolytic properties were investigated using linear cello- and xylooligomers as substrates. Cellobiose and xylobiose, respectively, were the final products (fig. 5.24D). Interestingly, a G3 and an X4 intermediate were formed from cellotetraose and xylopentaose, respectively, although in both cases no glucose could be detected (fig. 5.24D). This might be due to a possible transglycosidase activity of the enzyme under the experimental conditions used.

Such an activity is not uncommon to glycosidases that, like those grouped in family 51, hydrolyze their substrates by a retaining cleavage mechanism (Harjunp et al. (1999), Withers (2001)). The appearence of weak spots representing larger products than the starting substrates cello- or xylopentaose might be taken as further indication for the above notion (fig. 5.24D). Taken together, CelB has the hallmark qualities of an endoglucanase which acts mainly on CMC and cellulose but is also capable of hydrolyzing xylan.

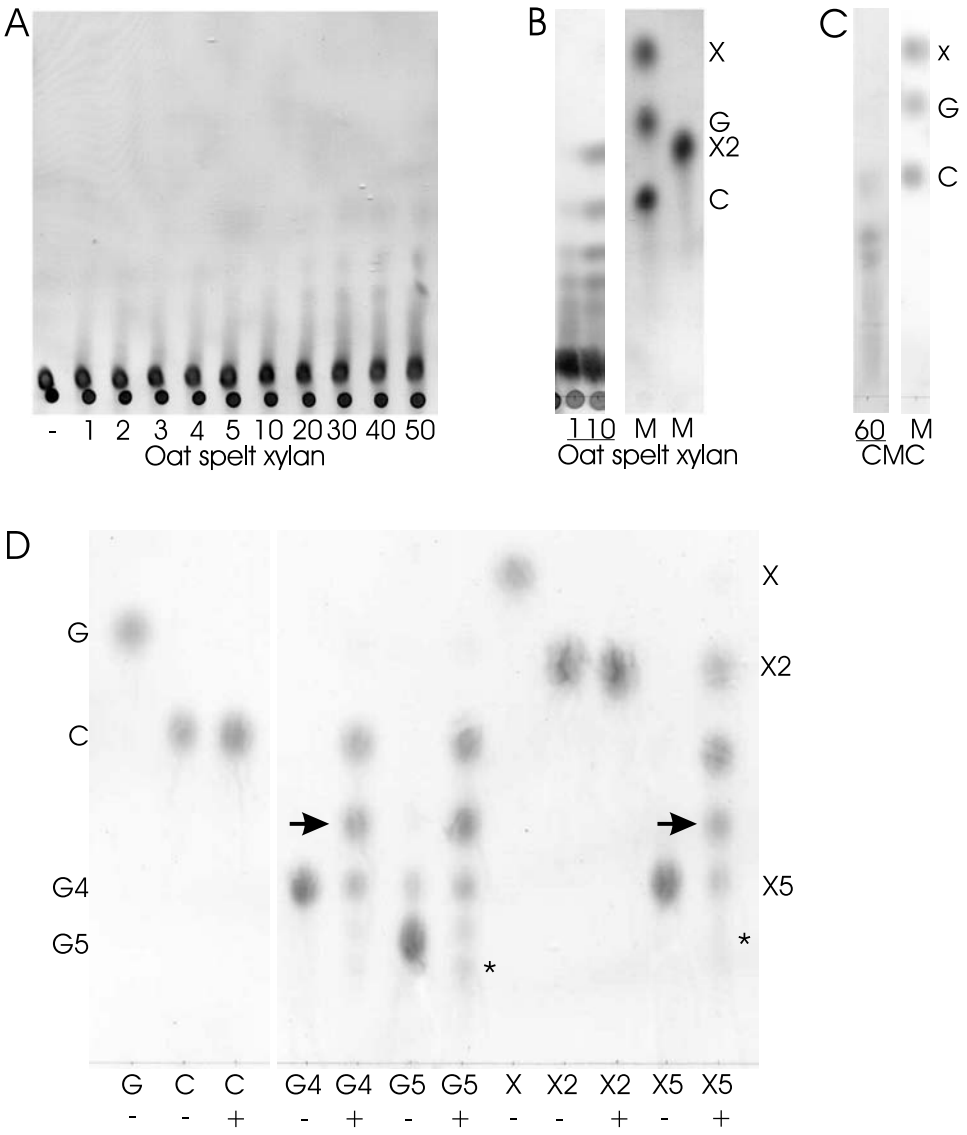


Figure 5.24: TLC analysis of degradation products of various substrates after incubation with $\mathbf{CelB_{trunc}}$. **(A)** Time course of degradation products of 0.25 % xylan at an enzyme concentration of 16 g ml^{-1}. **(B)** Hydrolysis of xylan at high enzyme concentrations (80 g ml^{-1}, *underlined*). **(C)** Extensive hydrolysis of 0.5 % CMC. **(D)** Hydrolysis (1 h) of oligosaccharides (10 mM) at an enzyme concentration of 16 g ml^{-1}. *Arrows* indicate G3 and X4 intermediates that might have arisen from degradation of larger transglycosylation products. Possible transglycosylation products larger than the starting substrates are marked by *asterisks*. Substrate blanks (-) were incubated without enzyme along with the samples with enzyme (+). *Numbers* give incubation time in minutes. *M*, marker; *G*, glucose; *C*, cellobiose; *G4*, cellotetraose; *G5*, cellopentaose; *X*, xylose; *X2* xylobiose.

6 Discussion

In the course of this work two glycosyl hydrolases from the thermoacidophile *A. acidocaldarius* were characterized. They displayed widely differing characteristics even though both showed highest activity on β-1,4-linked glucans. CelA was mostly active in the neutral pH range whereas CelB was an acidophilic enzyme. And while CelB released cellobiose and xylobiose from CMC and xylan, CelA produced only minor amounts of oligosaccharides from CMC even after prelonged incubation with the substrate.

The first approach towards elucidating xylan breakdown in *A. acidocaldarius* was based on expression screening of a plasmid gene bank. Even though the xylan-containing top agar contained a buffer at pH 3.5 the degrading activity observed in the clone containing the *celA* gene was later found to be due to an enzyme inactive below pH 4. The lysis zone was put down to insufficient buffering, probably because incubation over night and the nutrient agar at pH 7 led to an increase of pH in the top agar.

The *celA* gene encodes a CMCase belonging to family 9, subgroup E1 of glycoside hydrolases. The assignment is based on the highest sequence similarity to CelD of *C. thermocellum*, representative of the E1 subgroup, with respect to (*i*) presence of an Ig-like domain; (*ii*) absence of C-terminal CBD domains; and (*iii*) presence of the zinc binding site. In contrast, these features do not match those of the endoglucanase E4 of *T. fusca* that is a representative of the E2 subgroup.

One zinc and two calcium ions remained bound to CelA after extensive dialysis. Both CelD and E4 have divalent metal ions bound in their structures (Juy et al. (1992), Sakon et al. (1997)). For CelA the sequence alignment showed conservation of amino acids whose side chains interact with ions at the zinc binding site and the calcium binding site near the catalytic cleft (Zn^{2+}, $Ca^{2+}2$ in fig. 5.6B, p. 52). The second calcium binding site was not inferrable from the sequence, and may or may not coincide with one of the other two calcium binding sites found in CelD. Addition of EDTA to the enzyme resulted in a dramatic loss of stability. In accordance with results reported for CelD from *C. thermocellum* (Beguin and Alzari (1998)), this may be caused by removal of Zn^{2+} or Ca^{2+} from the respective binding sites.

The enzyme did not hydrolyze substrates that lack β-1,4-linkages, such as starch (α-1,4) and laminarin (β-1,3, β-1,6). It showed a marked preference for glucans over xylans (tab. 5.2, p. 56). This specificity is similar to that of Ced1 from *B. fibrisolvens* to which CelA aligns best (fig. 5.6B, p. 52, Berger et al. (1990)). In line with the activity of the enzyme on these substrates CelA expression was enhanced when cells were grown on CMC or oat spelt xylan. The enzyme exhibited a temperature optimum of 70 C which is slightly higher than the optimal growth temperature of the bacterium. In contrast, no activity and very low stability was monitored at pH 3.5 at which *A. acidocaldarius* grows best. This discrepancy may be explained by assuming an intracellular localization of the protein, since bacteria, including acidophiles keep a near neutral cytoplasmic pH (Bakker (1990)). This notion is supported by the lack of a signal sequence for secretion across the cytoplasmic membrane.

The reduction of CMC viscosity points towards an endocellulase-type of activity. On the other hand, CelA has a preference for producing short oligosaccharides from lichenan and released only small amounts of oligosaccharides from CMC even after prolonged incubation. However, the enzyme hydrolyzed small oligosaccharides, such as cellobiose, cellotriose, cellotetraose and aryl glycosides. The release of glucose from cellobiose is particularly interesting as Ced1 of *B. fibrisolvens* lacks this activity (Berger et al. (1990)). Most but not all family 9 enzymes, EGC from *F. succinogenes* being an exception (Bra-Maillet et al. (2000)), hydrolyze pNPC, while longer aryl glycosides, as in this study, were not tested. The hydrolysis of short soluble oligosaccharidic substrates by CelA points to a role of the enzyme in catabolizing short chain degradation products in the cytoplasm that are imported from the medium. The absence of a signal sequence is in line with this notion as is the lack of a CBD, since crystalline cellulose need not be degraded. A similar cytoplasmic function has also been proposed for a cyclomaltodextrinase from *A. acidocaldarius* (Matzke et al. (2000)).

Strikingly, CelA released pNP from pNPC and pNPG4, but not from pNPG and pNPG3, suggesting that the enzyme prefers to accommodate two glucosyl units on the minus side of the substrate binding groove rather than four. The activity seen on pNPG4 may actually be the result of initial degradation to cellobiose and pNPC, and subsequent breakdown of pNPC. This mode of binding is consistent with a lack of pNP release when digesting pNPG3 with CelA, which can nonetheless be shown to break down into cellobiose and pNPG. Possibly, such a digestion pattern could be the result of two strong binding subsites 1 and 2 or, alternatively, might be caused by steric hindrance of binding of β-1,4-linked glucans on the non-reducing side of these subsites. Fig. 6.1 shows this interpretation of the results. The determination of the CelA tertiary structure would certainly help in understanding this unusual pattern of activity.

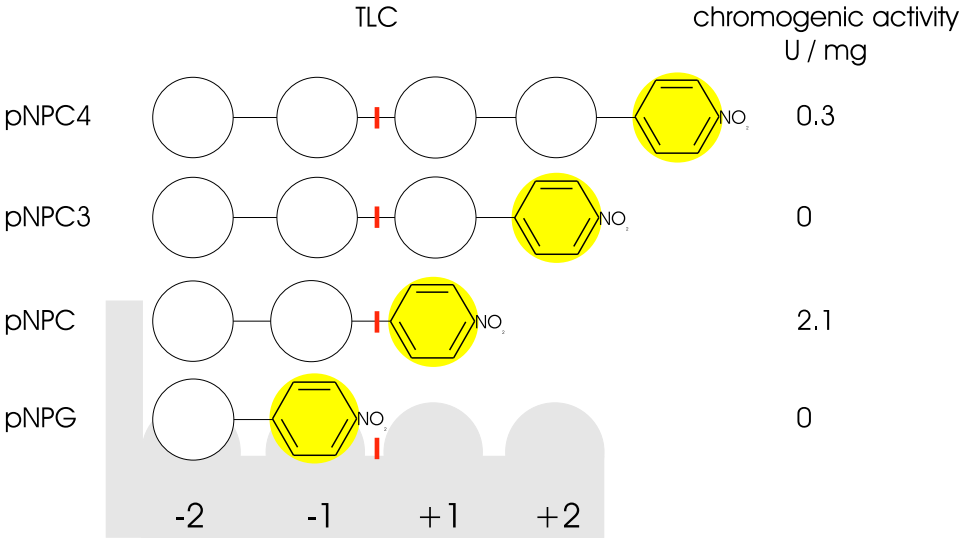


Figure 6.1: Schematic diagram of the catalytic site of CelA as deduced from activity against aryl glycosides. The pNP derivatives are represented by *circles* for the glucose units with chromogenic pNP substituents in *yellow*. The enzyme is in *grey* with subsites *numbered* according to Davies et al. (1997)). The *red bar* shows the site of hydrolysis as deduced by TLC. On the *right* chromogenic activity against the apropriate substrate is given.

Crystallization trials with CelA yielded crystals with ethylene glycol and MPD as precipitant. However crystals were generally needle-like and only in MPD was it possible to gain crystals of sufficient quality for data collection. Although analysis of the *celB* gene indicated that the tertiary structure was probably similar to that of CelD from *C. thermocellum* for which the three dimensional structure is known, molecular replacement trials did not yield a satisfactory model of the CelA structure.

Nonetheless the CelA crystals had some interesting properties. They did not dissolve in water and they did not dye with Izit. These two criteria are often used to distinguish microcrystalline precipitate from amorphous precipitate or denatured protein. The CelA crystals in MPD clearly show that these criteria are not absolute. MPD seems to be a competitive solute for Izit, which makes the dye accumulate in the solvent phase instead of binding to protein. Dissolving of the crystals in water must be energetically favourable, since the protein does not crystallize spontaneously in water, but the resilience of the crystals indicate that the crystal contacts pose a kinetic barrier to solvation in water. Hence streak seeding of precipitates obtained from initial crystal screens is a further, albeit labourious screening method for crystallization.

In the beginning of this work a 100 kDa protein had been noted that was expressed when cells were grown on CMC and xylan and was shown by zymogram analysis to be active against these polysaccharides. Interestingly it remained cell-associated during all stages of growth in batch culture, but it could be extracted with Triton X-100 from *A. acidocaldarius*. Purification and characterization of both wild-type and recombinant forms of the protein demonstrated it to be a thermoacidophilic endoglucanase, with activities against CMC, acid-swollen cellulase and oat spelt xylan.

The protein is remarkably stable at acidic pH and temperatures up to 80 C. These thermoacidophilic properties are in line with the growth characteristics of the organism which reflect its natural habitat. How tolerance to acidic pH is achieved in proteins is poorly understood. Schwermann et al. (1994) observed that acidophilic proteins possess a reduced density of both positively and negatively charged residues at their surface and proposed that this phenomenon might contribute to acidostability by preventing electrostatic repulsion at low pH. To test this hypothesis, the amino acid composition of the catalytic domain of CelB was compared with those from two other members of GH family 51, EGF from *F. succinogenes* and AbjA (Genpept acc. no. CAA76421) from the thermophile *Thermobacillus xylanilyticus* (fig. 6.2). EGF has a pH and temperature optimum of 5.3 and 40 C, respectively (Malburg et al. (1997)). For AbjA, the respective values are pH 5.9 and 75 C (Debeche et al. (2000)). CelB displayed a lower per-

centage of charged amino acids, especially lysine and arginine, which were reduced by together 10.1 % compared to EGF and 6.4 % in comparison with AbjA. On the other hand, CelB contains a higher percentage of alanine and proline and of uncharged polar residues. Thus, these data are at least not in contradiction to the above notion.

Wild-type and full-length recombinant CelB were found to be soluble only in the presence of detergent, while the truncated form of the protein ($CelB_{trunc}$), lacking the C-terminal 203 residues was readily soluble in buffer. A hydropathy plot of the protein is consistent with this observation, predicting a C-terminal hydrophobic region encompassing residues 700–900 (fig. 6.3). Furthermore, $CelB_{trunc}$ exhibited the same pH and temperature dependence as the wild-type protein, suggesting that the carboxy-terminal fragment is not essential for catalysis. Rather, these data point to a role in cell association, possibly by specific protein-protein interaction with the S-layer of the organism (Matzke (1999)), although sequence analysis did not identify a typical S-layer binding domain (Matuschek et al. (1996)). A possible approach to prove this notion would be to determine the cellular localization of homologously expressed C-terminally truncated variants of CelB. Unfortunately, such experiments are not feasible due to the fact that *A. acidocaldarius* cannot yet be manipulated by genetic means. However, some evidence in support of the above notion is provided by a study on an α-amylase (AmyA, Koivula et al. (1993)) from the same organism. Like CelB, AmyA remains attached to the cells during exponential growth (Schwermann et al. (1994)), but it is released into the medium as the culture approaches the stationary phase. And like CelB, the cell-associated form of the enzyme is extractable by Triton X-100. Interestingly, a hydropathy plot of AmyA revealed a hydrophobic N-terminal region ($\sim$ residues 110–340) (fig. 6.3) to which a function has not yet been assigned. Sequence comparison of this portion with the hydrophobic region of CelB revealed similarities over a stretch of 11 amino acids (fig. 6.3B) with the consensus sequence:

```
          12
PG(D,Q)TVThSXXhA
```

Where 'X' denotes uncertain positions and 'h' indicates hydrophobic positions. This short motif appeared twice in CelB separated by 91 amino acids. In AmyA, the first sequence is separated by 92 amino acids from the second region in which the motif is repeated (fig. 6.3, inset). Whether these motifs are of any significance for the function of the C-terminal portion of CelB remains to be established.

Between the catalytic domain and the hydrophobic region a stretch of 20 amino acids (residues Ser 720–Asp 739) was found with 60 % of the residues being proline, aspartate, serine or threonine which are typical of linker sequences (Gilkes et al. (1991)):

```
CelB: 720 SSPTVTAGGSETVTASFSSD 739
```

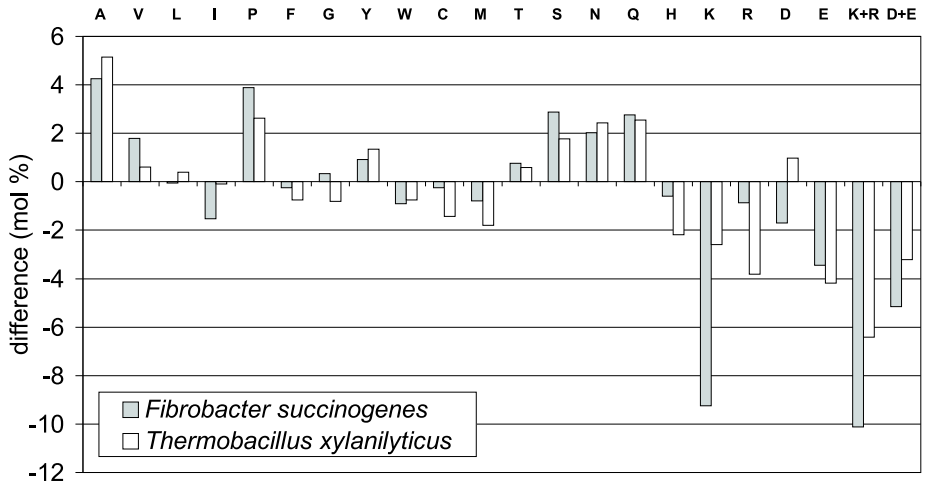


Figure 6.2: Comparison of the amino acid composition of the catalytic domain of CelB from *A. acidocaldarius* with CelF from *F. succinogenes* and AbjA from *T. xylanilyticus*. Given is the difference in mol % for CelB with regard to the other proteins. Positive values indicate that the catalytic domain of CelB has a higher percentage of a certain amino acid than the catalytic domain of the indicated protein.

This was the highest occurrence of these amino acids in the whole sequence. Thus, this region may function as a flexible linker between the catalytic domain and the C-terminal portion attached to the cell.

High activity of CelB against CMC and TLC analysis of the time course with initial production of high molecular mass products from CMC and oat spelt xylan characterize the enzyme as an endoglucanase. In terms of substrate specificity CelB is similar only to one other member of GH family 51, an endoglucanase (EGF) from *Fibrobacter succinogenes*. Like CelB, EGF has no activity on pNPAraf. This is also in contrast to all other members of this family for which such an activity was tested (Beylot et al. (2001), Debeche et al. (2000), Kim et al. (1998), Lee et al. (2001), Matsuo et al. (2000)). Only in the case of AbfA from *Geobacillus stearothermophilus* T-6 (Gilead and Shoham (1995)), a very low activity on CMC (0.08 % of the arabinofuranosidase activity) was reported. This remarkable finding is further corroborated by the fact that, based on sequence comparison of their catalytic domains, CelB and EGF form a distinct cluster within family 51, although both enzymes differ in their pH and temperature optima. Thus, substrate specificity seems to put more constraints on the sequence than do pH and temperature adaptation. EGF and CelB are examples of divergent evolution in a family of proteins which results in new substrate specificity.

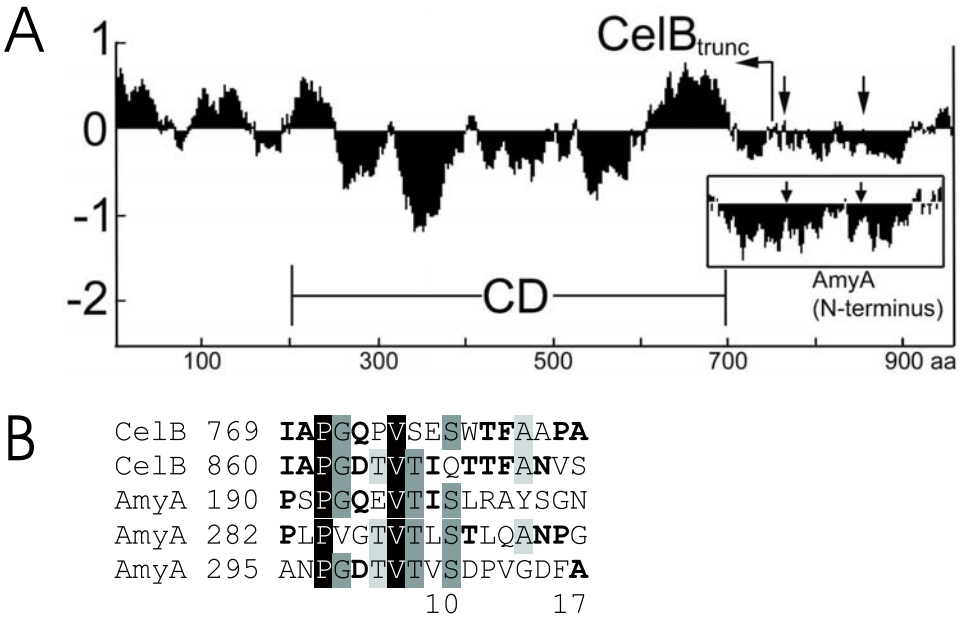


Figure 6.3: (A) Hydropathy plot of CelB. *Horizontal arrow* indicates the C-terminus of CelB$_{\text{trunc}}$. The catalytic domain (*CD*) is marked. The *inset* shows a hydropathy plot of the hydrophobic region of AmyA ($\sim$ residues 110–340) from *A. acidocaldarius* drawn to the same scale. *Vertical arrows* show the position of conserved sequence motifs detailed in (B). **(B) Sequence alignment of a repeated motif found in the hydrophobic regions of CelB and AmyA from *A. acidocaldarius*.** The *numbers* give the positions in the sequence of the preprotein. Amino acids identical in five, four, three or two sequences are shaded *black*, *gray*, *light gray* or *bold-face*, respectively.

Oat spelt xylan, birchwood xylan, cellobiose and CMC as sole carbon source result in expression of the *celB* gene. Whereas oat spelt xylan and CMC are readily degraded by CelB, birchwood xylan is not. Lack of activity against birchwood xylan may be due to its high level of acetylation (Coughlan and Hazlewood (1993)) which may cause steric hindrance resulting in low activity of CelB against this substrate even though it leads to gene expression. Because *celB* is not expressed during growth on various other substrates a specific induction rather than lack of repression seems responsible for expression. Which part of the substrate molecules are capable of induction is unclear. Cellobiose induces expression and in the presence of CMC low basal levels of CelB or acid hydrolysis due to the acidic surroundings could well lead to the formation of this small inducer molecule. It cannot be said whether the inducer molecule is first transported into the cell and then interacts with a transcriptional regulator or if it is recognized on the cell surface. CelB is not part of an operon, but is surrounded by no less than three transcriptional regulators and also a histidine kinase which may be part of a two component regulatory system. Moreover 3 kb downstream of *celB* lies a region displaying homologies to genes of a binding protein-dependent (ABC) transport system and upstream of CelB is a putative gene which may code for a permease. This leaves both mechanisms open to discussion. The palindrome found just downstream of the promotor may be an operator for a transcriptional repressor (fig. 5.17B, p. 72). As far as induction by xylan is concerned regulation could take place by a different signalling pathway or by the same which does not discriminate between similar inducer molecules. The presence of a possible CRE downstream of the *celB* promotor presents an additional route by which regulation may take place. In *Bacillus subtilis* and other low-GC gram-positive bacteria CREs play a central role in global regulation of catabolite repression. CREs are recognized by a protein complex activated by phosphorylation in the presence of glucose. DNA binding, mediated by CcpA in *B. subtilis*, leads to repression of gene expression (Stlke and Hillen (1999)). If such a regulatory mechanism exists for *celB* or even *A. acidocaldarius* cannot be said at present, but the possibility of catabolite repression via CREs in this high-GC gram-positive bacterium is intriguing.

Induction and activity suggest that CelB may play a role in the degradation of both cellulose and xylan *in vivo*. This is further underlined by the fact that hydrolysis of xylan and CMC by CelB leads to formation of disaccharides, albeit at high enzyme concentrations. Complete degradation of the substrates is likely to be performed more effectively and completely by cooperative efforts of CelB with other glycoside hydrolases. Indeed, zymogram analysis demonstrated the presence of additional protein bands with CMCase and xylanase activity in the Triton X-100 extract (fig. 5.14, p. 68). The resulting oligosaccharides may then be transported into the cytoplasm. Transport could be mediated by a binding protein-dependent transport com-

plex. Such a complex has been found for maltose and maltodextrins in *A. acidocaldarius* (Hlsmann et al. (2000a), b) and transporters of the same kind have been shown to mediate cello- and xylooligosaccharide uptake in other gram positive bacteria (Schlsser et al. (1999), Shulami et al. (1999)). From the sequence data alone it cannot be said if the ABC transporter encoded downstream of the *celB* gene plays such a role. Once inside the cell, oligosaccharides may be hydrolyzed by enzymes such as CelA, leading to formation of disaccharides. Together CelA and CelB may thus be responsible for a large part of the degradation of cellulose and at least main chain hydrolysis of certain types of xylan. Cellobiose and xylobiose could then be hydrolyzed by β-glycosidases. The resulting xylose and glucose can be further metabolized by entry into the pentose phosphate cycle and glycolysis, respectively. The overall scheme is depicted in fig. 6.4.

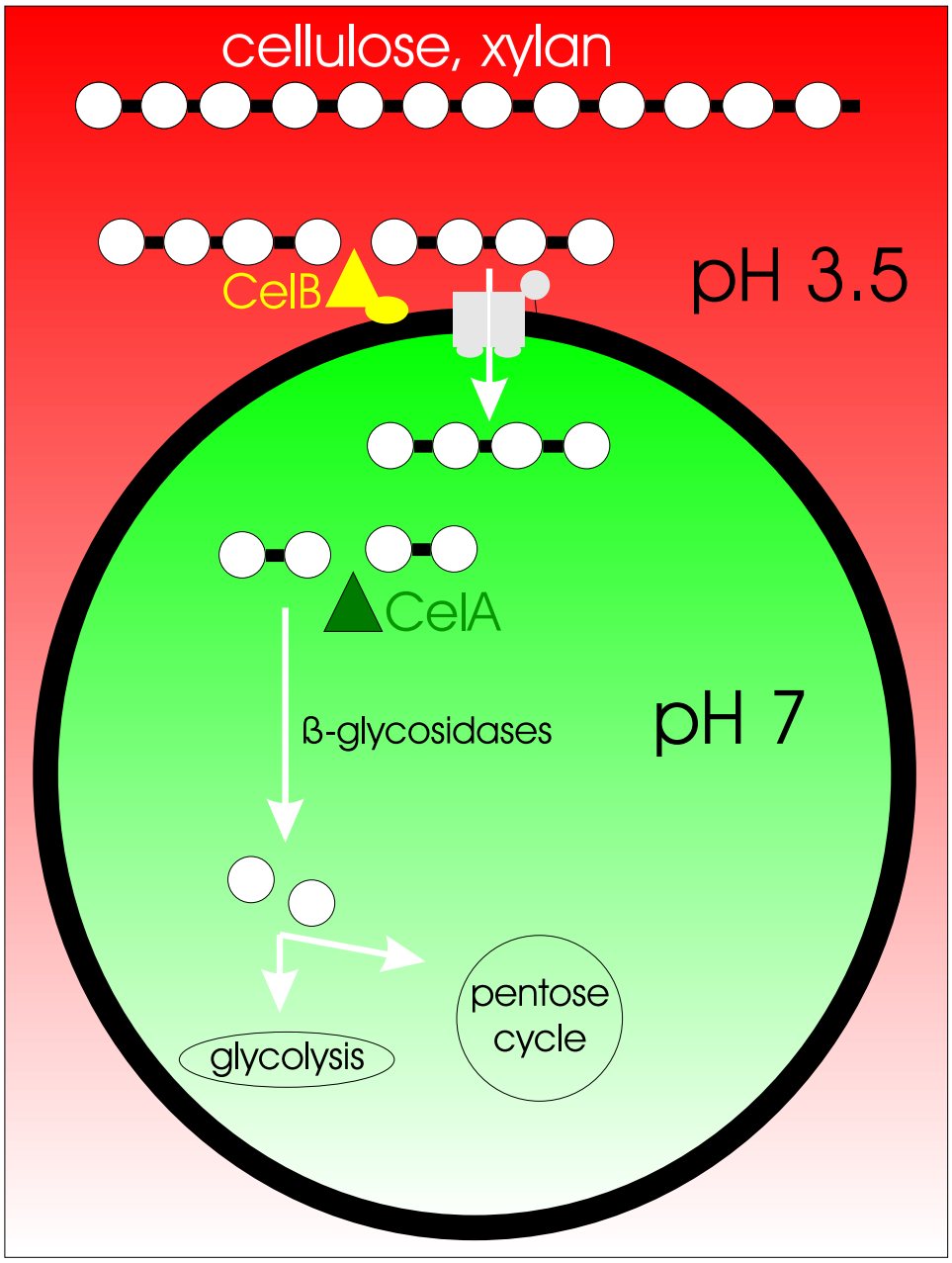


Figure 6.4: Scheme of the possible roles of CelA and CelB in cellulose and xylan degradation. CelB is cell-associated but facing the outside medium and may degrade extracellular polysaccharides. After these have been imported into the cell by a putative binding protein-dependent transport system CelA degrades the oligosaccharides to disaccharides. Cellobiose and xylobiose are then hydrolyzed by β-glycosidases. Catabolism proceeds via glycolysis or the pentose phosphate pathway.

Literature

Aguilar, C.F. and I. Sanderson, and M. Moracci, and M. Ciaramella, and R. Nucci, and M. Rossi, and L.H. Pearl (1997). Crystal structure of the β-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: Resilience as a key factor in thermostability. *J. Mol. Biol.*, 271:789–802.

Altschul, S.F. and Madden, T.L. and Schffer, A.A. and Zhang, J. and Zhang, Z. and Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–3402.

Bacon, J.S.D. and Gordon, A.H. and Morris, E.J. (1975). Acetyl groups in cell-wall preparations from higher plants. *Biochem. J.*, 149:485–487.

Bakker, E.P. (1990). The role of alkali-cation transport in energy coupling of neutrophilic and acidophilic bacteria: an assesment of methods and concepts . *FEMS Microbiol. Rev.*, 75:319–334.

Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.*, 6:37–40.

Beguin, P. (1994). The biological degradation of cellulose. *FEMS Microbiol. Rev.*, 13:25–58.

Beguin, P. and Alzari, P.M. (1998). The cellulosome of *Clostridium thermocellum*. *Biochem. Soc. Trans.*, 26:178–184.

Berger, E. and Jones, W.A. and Jones, D.T. and Woods, D.R. (1990). Sequencing and expression of a cellodextrinase (*ced1*) gene from *Butyrivibrio fibrisolvens* H17c cloned in *Escherichia coli*. *Mol. Gen. Genet.*, 223:310–318.

Bergquist, P.L. and M.D. Gibbs and V.S.J. Teo and D.J. Saul and H.W. Morgan (1999). Molecular diversity of thermophilic cellulolytic and hemicellulolytic bacteria. *FEMS Microbiol. Ecol.*, 28:99–110.

Beylot, M. and McKie, V.A. and Voragen, A.G.J. and Doeswijk-Voragen, C.H.L. and Gilbert, H.J. (2001). The *Pseudomonas cellulosa* glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity. *Biochem. J.*, 358:607–614.

Blanco, A. and Vidal, T. and Colom, J.F. and Pastor, F.I.J. (1995). Purification and properties of Xylanase A from alkali-tolerant *Bacillus* sp. strain BP-23. *Appl. Environ. Microbiol.*, 61:4468–4470.

Blum, H. and Beier, H. and Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, 8:93–99.

Bourne, Y. and Henrissat, B. (2001). Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.*, 11:593–600.

Bra-Maillet, C. and Broussolle, V. and Pristas, P. and Girardeau, J. and Guadet, G. and Forano, E. (2000). Characterisation of endoglucanases EGB and EGC from *Fibrobacter succinogenes* S85. *Biochim. Biophys. Acta*, 1476:191–202.

Brett, C.T. and Waldron, K.W. (1996). Physiology and biochemistry of plant cell walls. *London: Chapman and Hall.*

Cann, I.K.O. and Kocherginskaya, S. and King, M.R. and White, B.A. and Mackie, R.I. (1999). Molecular Cloning, sequencing, and expression of a novel multidomain mannanase gene from *Thermoanaerobacterium polysaccharolyticum*. *J. Bacteriol.*, 181:1643–1651.

Charnock, S.J. and Spurway, T.D. and Xie, H. and Beylot, M. and Virden, R. and Warren, R. and Hazlewood, G.P. and Gilbert, H.J. (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *J. Biol. Chem.*, 273:32187–32199.

Chayen, N.E. (1997). A novel technique to control the rate of vapour diffusion giving larger protein crystals. *J. Appl. Cryst.*, 30:198–202.

Chen, P. and Buller, S. (1995). Activity staining of xylanases in polyacrylamide gels containing xylan. *Anal. Biochem.*, 226:186–188.

Clarke, J.H. and Davidson, K. and Gilbert, H.J. and Fontes, C.M.G.A. and Hazlewood, G.P. (1996). A modular xylanase from mesophilic *Cellulomonas fimi* contains the same cellulose-binding and thermostabilizing domains as xylanases from thermophilic bacteria. *FEMS Microbiol. Lett*, 139:27–35.

Coughlan, M.P. and Hazlewood, G.P. (1993). β-1,4-D-Xylan-degrading enzyme systems: biochemistry, molecular biology and applications. *Biotechnol. Appl. Biochem.*, 17:259–289.

Coutinho, P.M. and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. *Cambridge: The royal society of cambridge.*

Cuff, J.A. and Clamp, M. and Siddiqui, A.S. and Finlay, M. and Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics*, 14:892–893.

Daniel, R.M. and Dines, M. and Petach, H.H. (1996). The denaturation and degradation of stable enzymes at high temperatures. *Biochem. J.*, 317:1–11.

Darland, G. and Brock, T.D. (1971). *Bacillus acidocaldarius sp. nov.*, an acidophilic thermophilic spore-forming bacterium. *J. Gen. Microbiol.*, 67:9–15.

Davies, G.J. and Wilson, K.S. and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.*, 321:557–559.

Debeche, T. and Cummings, N. and Connertion, I. and Debeire, P. and O' Donohue, M.J. (2000). Genetic and biochemical characterization of a highly thermostable α-L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Appl. Environ. Microbiol.*, 66:1734–1736.

Decker, E. (1987). Isolierung und Charakterisierung von Plasmiden aus thermoacidophilen Bacilli. Diploma thesis. *Tbingen: University of Tbingen, Germany.*

Deinhard, G. and Poralla, K. (1996). Vorkommen, Biosynthese und Funktion alicyclischer Fettsuren bei Bakterien. *Biospektrum*, 2:40–46.

Deshpande, M.V. and Erikson, R.E. and Petterson, L.G. (1984). An assay for selective determination of exo-1,4-β-glucanases in a mixture of cellulolytic enzymes. *Anal. Biochem.*, 138:481–487.

Dower, D.J. (1988). High efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Res.*, 16:6127–6145.

Dsterhft, E. and Linssen, V.A.J.M. and Voragen, A.G.J. and Beldman, G. (1997). Purification, characterization, and properties of two xylanases from *Humicola insolens*. *Enzyme Microb. Technol.*, 20:437–445.

Eckert, K. (1998). Biochemische Charakterisierung einer xylan-abbauenden Aktivitt aus *Alicyclobacillus acidocaldarius* ATCC27009, Diploma thesis. *Berlin: Humboldt-Universitt zu Berlin.*

Eckert, K. and Ernst, H.A. and Schneider, E. and Larsen, S. and Lo Leggio, L. (2003). Crystallization and preliminary X-ray analysis of *Alicyclobacillus acidocaldarius* endoglucanase CelA. *Acta Crystallogr. D. Biol. Crystallogr.*, 59:139–141.

Eisel, D. and Grnewald-Janho, S. and Hloch, P. and Hfner, P. and Keesey, J. and Kruchen, B. and Rger, B. (2000). DIG application manual for filter hybridization. *Germany: Roche.*

Engelhardt, H. and Peters, J. (1998). Structural research on surface layers: A focus on stability, surface layer homology domains, and surface layer-cell wall interactions. *J. Struct. Biol.*, 124:276–302.

Fernandes, A.C. and Fontes, C.M.G.A. and Gilbert, H.J. and Hazlewood, G.P. and Fernandes Tito, H. and Ferreira, L.M.A. (1999). Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem. J.*, 342:105–110.

Gilead, S. and Shoham, Y. (1995). Purification and characterization of α-L-arabinofuranosidase from *Bacillus stearothermophilus* T-6. *Appl. Environ. Microbiol.*, 61:170–174.

Gilkes, N.R. and Henrissat, B. and Kilburn, D.G. and Miller, R.C. and Warren, R.A.J. (1991). Domains in microbial β-1,4-glycanases: Sequence conservation, function and enzyme families. *Microbiol. Rev.*, 55:303–315.

Gill, J. and Rixon, J.E. and Bolam, D.N. and McQueen-Mason, S. and Simpson, P.J. and Williamson, M.P. and Hazlewood, G.P. and Gilbert, H.J. (1999). The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic acitivity against complex substrates by a common mechanism. *Biochem. J.*, 342:473–480.

Gruber, K. and Klintschar, G. and Hayn, M. and Schlachter, A. and Steiner, W. and Kratky, C. (1998). Thermophilic xylanase from *Thermomyces lanuginosus*: High-resolution X-ray structure and modeling studies. *Biochemistry*, 37:13475–13485.

Haney, P.J. and Badger, J.H. and Buldak, G.L. and Reich, C.I. and Woese, C.R. and Olsen, G.J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci.*, 96:3578–3583.

Harjunp, V. and Helin, J. and Koivula, A. and Siika-aho, M. and Drakenberg, T. (1999). A comparative study of two retaining enzymes of *Trichoderma reesei*: transglycosylation of oligosaccharides catalyzed by the cellobiohydrolase I, Cel7A, and β-mannanase, Man5A. *FEBS Lett.*, 443:149–153.

Harwood, C.R. and Cutting, S.M. (1990). Molecular biological methods for *Bacillus*. *London: John Wiley and sons.*

Hayashi, H. and Takagi, K. and Fukumura, M. and Kimura, T. and Karita, S. and Sakka, K. and Ohmiya, K. (1997). Sequence of *xynC* and properties of XynC, a major component of the *Clostridium thermocellum* cellulosome. *J. Bacteriol.*, 179:4246–4253.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, 280:309–316.

Henrissat, B. and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, 293:781–788.

Henrissat, B. and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.*, 316:695–696.

Henrissat, B., and Callebaut, I. and Fabrega, S. and Lehn, P., and Mornon, J. and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci.*, 92:7090–7094.

Henrissat, B. and Teeri, T.T. and Warren, R.A.J. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett.*, 425:352–354.

Hippchen, B. and Rll, A. and Poralla, K. (1981). Occurence in soil of thermoacidophilic bacilli possessing cyclohexane fatty acids and hopanoids. *Arch. Microbiol.*, 129:53–55.

Hiromi, K. and Ohnishi, M. and Tanaka, A. (1983). Subsite structure and ligand binding mechanism of glucoamylase. *Mol. Cell. Biochem.*, 51:79–95.

Hlsmann, A. (2000b). Identifizierung und Charakterisierung eines Bindeprotein-abhngigen Maltose-Transportsystems bei dem extremophilen grampositiven Bakterium *Alicyclobacillus acidocaldarius*. *PhD thesis. Berlin: Humboldt-Universitt zu Berlin.*

Hlsmann, A. and Lurz, R. and Scheffel, F. and Schneider, E. (2000a). Maltose and maltodextrin transport in the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius* is mediated by a high-affinity transport system that includes a maltose binding protein tolerant to low pH. *Journal of bacteriology*, 182:6292–6301.

Hner zu Bentrup, K. and Schmid, R. and Schneider, E. (1994). Maltose transport in *Aeromonas hydrophila*: purification, biochemical characterization and partial amino acid sequence analysis of a periplasmic maltose binding protein. *Microbiology*, 140:945–951.

Hrmova, M. and Fincher, G.B. (1993). Purification and properties of three (1,3)-β-D-glucanase isoenzymes from young leaves of barley (*Hordeum vulgare*). *Biochem. J.*, 289:453–461.

Huang, X. and Miller, W. (1981). A space-efficient algorithm for local similarities. *Adv. Appl. Math.*, 12:373–381.

Hueck, C.J. and Hillen, W. (1995). Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the gram-positive bacteria? *Mol. Microbiol.*, 15:395–401.

Inagaki, K. and Nakahira, K. and Kazuhisa, M. and Tamura, T. and Tanaka, H. (1998). Gene cloning and characterization of an acidic xylanase from *Acidobacterium capsulatum*. *Biosci. Biotechnol. Biochem.*, 62:1061–1067.

Juy, M. and Amit, A.G. and Alzari, P.M. and Poljak, R.J. and Claeyssens, M. and Beguin, P. and Aubert, J.P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature*, 357:89–91.

Kelly, L.A. and MacCallum, R.M. and Sternberg, M.J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, 299:499–520.

Kim, K.S. and Lilburn, T.G. and Renner, M.G. and Breznak, J.A. (1998). *arfI* and *arfII*, two genes encoding α-L-arabinofuranosidases in *Cytophaga xylanolytica*. *Appl. Environ. Microbiol.*, 64:1919–1923.

Kimura, T. and Ito, J. and Kawano, A. and Makino, T. and Kondo, H. and Karita, S. and Sakka, K. and Ohmiya, K. (2000). Purification, characterization, and molecular cloning of acidophilic xylanase from *Penicillium* sp. 40. *Biosci. Biotechnol. Biochem.*, 64:1230–1237.

Koivula, T.T. and Hemil, H. and Pakkanen, R. and Sibakov, M. and Palva, I. (1993). Cloning and sequencing of a gene encoding acidophilic amylase from *Bacillus acidocaldarius*. *J. Gen. Microbiol.*, 139:2399–2407.

Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.*, 28:416–436.

Krengel, U. and Dijkstra, B.W. (1996). Three-dimensional structure of endo-1,4-β-xylanase I from *Aspergillus niger*: Molecular basis for its low pH optimum. *J. Mol. Biol.*, 263:70–78.

Kulkarni, N. and Shendye, A. and Rao, M. (1999). Molecular and biotechnological aspects of xylanases. *FEMS Microbiol. Rev.*, 23:411–456.

Kumar, P.R. and Eswaramoorthy, S. and Vithayathil, P.J. and Viswamitra, M.A. (2000). The tertiary structure at 1.59 Å resolution and the proposed amino acid sequence of a family-11 xylanase from the thermophilic fungus *Paecilomyces varioti* Bainier. *J. Mol. Biol.*, 295:581–593.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157:105–132.

La, D. and Silver, M. and Edgar, R.C. and Livesay, D.R. (2003). Using motif-based methods in multiple genome analyses: A case study comparing orthologous mesophilic and thermophilic proteins. *Biochemistry*, 42:8988–8998.

Laemmli, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature*, 227:680–685.

Laurie, J.I. and Clarke, J.H. and Ciruela, A. and Faulds, C.B. and Williamson, G. and Gilbert, H.J. and Rixon, J.E., Millward-Sadler, J. and Hazlewood, G.P. (1997). The NodB domain of a multidomain xylanase from *Cellulomonas fimi* deacetylates acetylxylan. *FEMS Microbiol. Lett.*, 148:261–264.

Laursen, J. (1985). Biomolecular thin film samples analysed by EDXRF. *Colloquium Spectroscopicum Internationale XXIV*, pages 38–39.

Laursen, J. and Stikans, M. and Karlsen, K. and Pind, N. (1998). A versatile and easy to handle EDXRF instrumentation. *Proceedings of the european conference on energy dispersive X-ray spectrometry. Bologna: Editrice Compositori.*

Lawson, S.L. and Wakarchuk, W.W. and Withers, S.G. (1997). Positioning the acid/base catalyst in a glycosidase; studies with *Bacillus circulans* xylanase. *Biochemistry*, 36:2257–2265.

Lee, R.C. and Burton, R.A. and Hrmova, M. and Fincher, G.B. (2001). Barley arabinoxylan arabinofuranohydrolases: purification, characterization and determination of primary structures from cDNA clones. *Biochem. J.*, 356:181–189.

Lee, Y. and Lowe, S.E. and Henrissat, B. and Zeikus, G. (1993). Characterisation of the active site and thermostability regions of endoxylanase from *Thermoanaerobacterium Saccharolyticum* B6A-RI. *J. Bacteriol.*, 175:5890–5898.

Liu, S. and Wiegel, J. and Gherardini, F.C. (1996). Purification and cloning of a thermostable xylose (glucose) isomerase with an acidic pH optimum from *Thermoanaerobacterium* strain JW/SL-YS 489. *J. Bacteriol.*, 178:5938–5945.

Lopez-Fernandez, C.L. and Rodriguez, J. and Ball, A.S. and Copa-Patino, J.L. and Perez-Leblic, M.I. and Arias, M.E. (1998). Application of the affinity binding of xylanases to oat-spelt xylan in the purification of endoxylanase CM-2 from *Streptomyces chattanoogensis* CECT 3336. *Appl. Microbiol. Biotechnol.*, 199850:284–287.

Malburg, S.R.C. and Malburg, L.M.J. and Liu, T. and Iyo, A.H. and Forsberg, C.W. (1997). Catalytic properties of the cellulose-binding Endoglucanase F from *Fibrobacter succinogenes* S85. *Appl. Environ. Microbiol.*, 63:2449–2453.

Matsuo, N.K.S. and Kuno, A. and Kobayashi, H. and Kusakabe, I. (2000). Purification, characterization and gene cloning of two α-L-arabinofuranosidases from *Streptomyces chartreusis* GS901. *Biochem. J.*, 346:9–15.

Matuschek, M. and Sahm, K. and Zibat, A. and Bahl, H. (1996). Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolasees with conserved S-layer-like domains. *Mol. Gen. Genet.*, 252:493–496.

Matzke, J. (1999). Untersuchungen zur Struktur und Acidophilie einer α-Amylase und Charakterisierung einer Neopullulanase (Cyclomaltodextrinase) aus *Alicyclobacillus acidocaldarius* ATCC 27009. PhD Thesis. *Osnabrck: University of Osnabrck*.

Matzke, J. and Herrmann, A. and Schneider, E. and Bakker, E.P. (2000). Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmatic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes. *FEMS Microbiol. Lett.*, 183:55–61.

Meissner, H. and Liebl, W. (1998). *Thermotoga maritima* maltosyltransferase, a novel type of maltodextrin glycosyltransferase acting on starch and maltooligosaccharides. *Eur. J. Biochem.*, 250:1050–1058.

Meissner, K. and Wassenberg, D. and Liebl, W. (2000). The thermostabilizing domain of the modular xylanase XynA of *Thermotoga maritima* represents a novel type of binding domain with affinity for soluble xylan and mixed-linkage β-1,3/β-1,4-glucan. *Mol. Microbiol.*, 36:898–912.

Messner, P. (1994). Personal communication. *Zentrum fr Ultrastrukturforschung Universitt fr Bodenkultur Wien*.

Millward-Sadler, S.J. and Davidson, K. and Hazlewood, G.P. and Black, G.W. and Gilbert, H.J. and Clarke, J.H. (1995). Novel cellulose-binding domains NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellvibrio mixtus*. *Biochem. J.*, 312:39–48.

Morelle, G. (1989). A plasmid extraction procedure on a miniprep scale. *Focus (BRL)*, 11:7–8.

Natesh, R. and Bhanumoorthy, P. and Vithayathil, P.J. and Sekar, K. and Ramakumar, S. and Viswamitra, M.A. (1999). Crystal Structure at 1.8 Å resolution and proposed amino acid sequence of a thermostable xylanase from *Thermoascus aurantiacus*. *J. Mol. Biol.*, 288:999–1012.

Neidhardt, F.C. and Ingraham, J.L. and Schaechter, M. (1990). Physiology of the bacterial cell. *Massachusetts: Sinauer Associates*.

Nicholas, K.B. and Nicholas, H.B. Jr. and Deerfield, D.W. (1997). GeneDoc: analysis and visualisation of genetic variation. *EMBNEW.NEWS*, 4.

Nielsen, H. and Engelbrecht, J., Brunak, S. and von Heihne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, 10:1–6.

Ohta, K. and Moriyama, S. and Tanaka, H. and Shige, T. and Akimoto, H. (2001). Purification and characterization of an acidophilic xylanase from *Aureobasidium pullulans* var. *melanigenum* and sequence analysis of the encoding gene. *J. Biosci. Bioeng.*, 92:262–270.

Page, R. (1996). An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, 12:357–358.

Park, K.H. and Kim,M.J. and Lee, H.S. and Han, N.S. and Kim, D. and Robyt, J.F. (1998). Transglycosylation reactions of *Bacillus stearothermophilus* maltogenic amylase with acarbose and various acceptors. *Carbohydr. Res.*, 313:235–246.

Przylas, I. and Tomoo, K. and Terada, Y. and Takaha, T. and Fujii, K. and Saenger, W. and Strter, N. (2000). Crystal structure of amylomaltase from *Thermus aquaticus*, a glycosyltransferase catalysing the production of large cyclic glucans. *J. Mol. Biol.*, 296:873–886.

Rice, D.W. and Yip, K.S.P. and Stillman, T.J. and Britton, K.L. and Fuentes, A. and Connerton, I. and Pasquo, A. and Scandurra, R. and Engel, P.C. (1996). Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol. Rev.*, 18:105–117.

Rouvinen, J. and Bergfors, T. and Teeri, T. and Knowles, J.K.C. and Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*, 249:380–386.

Ruiz, M. and Di Pietro, A. and Isabel, M. and Roncero, G. (1997). Purification and characterisation of an acidic endo-β-1,4-xylanase from the tomato vascular pathogen *Fusarium oxysporum* f. sp. *lycopersici*. *FEMS Microbiol. Lett.*, 148:75–82.

Sakon, J. and Adney, W.S. and Himmel, M.E. and Thomas, S.R. and Karplus, A. (1996). Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry*, 35:10648–10660.

Sakon, J. and Irwin, D. and Wilson, D.B. and Karplus, P.A. (1997). Structure and mechanism of endo / exocellulase E4 from *Thermomonospora fusca*. *Nat. Struc. Biol.*, 4:810–818.

Saloheimo, M. and Kuja-Panula, J. and Ylsmki, E. and Ward, M. and Penttil, M. (2002). Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (Cel1A). *Appl. Environ. Microbiol.*, 68:4546–4553.

Sambrook, J. and Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.*

Sara, M. (2001). Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? *Trends Microbiol.*, 9:47–49.

Schlsser, A. and Jantos, J. and Hackmann, K. and Schrempf, H. (1999). Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl. Environ. Microbiol.*, 65:2636–2643.

Schmidt, A. and Gbitz, G.M. and Kratky, C. (1999). Xylan binding subsite mapping in the xylanase from *Penicillium simplicissimum* using xylooligosaccharides as cryo-protectant. *Biochemistry*, 38:2403–2412.

Schmidt, A. and Schlacher, A. and Steiner, W. and Schwab, H. and Kratky, C. (1998). Structure of the xylanase from *Penicillium simplicissimum*. *Protein Sci.*, 7:2081–2088.

Schneider, E. and Wilken, S. and Schmid, R. (1994). Nucleotide-induced conformational changes of MalK, a bacterial ATP binding cassette transporter protein. *J. Biol. Chem.*, 269:20456–20461.

Schwarz, W.H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria . *Appl. Microbiol. Biotechnol.*, 56:634–649.

Schwermann, B. and Pfau, C. and Liliensiek, B. and Schleyer, M. and Fischer, T. and Bakker, E.P. (1994). Purification, properties and structural aspects of a thermoacidophilic α-Amylase from *Alicyclobacillus acidocaldarius* atcc 27009, insight into acidostability of proteins. *Eur. J. Biochem.*, 226:981–991.

Shulami, S. and Gat, O. and Sonenshein, A.L. and Shoham, Y. (1999). The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J. Bacteriol.*, 181:3695–3704.

Sinnott, M.L. (1991). Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.*, 90:1170–1202.

Sitte, P. and Ziegler, H. and Ehrendorfer, F. and Bresinsky, A. (1991). Strasburger Lehrbuch der Botanik. *Stuttgart: Gustav Fischer Verlag.*

Sleytr, U.B., and Sra, M. (1997). Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. *TIBTECH*, 15:20–26.

Somogyi, M. (1951). Notes on sugar determination. *Journal of biological chemistry*, 272:17523–17530.

Stlke, J. and Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.*, 2:195–201.

Stura, E.A. and Nemerow, G.R. and Wilson, I.A. (1992). Strategies in the crystallization of glycoproteins and protein complexes. *J. Cryst. Growth*, 122:273–285.

Sun, J.L. and Sakka, K. and Karita, S. and Kimura, T. and Ohmiya, K. (1998). Adsorption of *Clostridium stercorarium* Xylanase A to insoluble xylan and the importance of the CBDs to xylan hydrolysis. *J. Ferment. Bioeng.*, 85:63–68.

Takada, G. and Karita, S. and Takeuchi, A. and Ahsan, M.M. and Kimura, T. and Sakka, K. and Ohmiya, K. (1996). Specific adsorption of *Clostridium stercorarium* xylanase to amorphous cellulose and its desorption by cellobiose. *Biosci. Biotechnol. Biochem.*, 60:1183–1185.

Teather, R.M. and Wood, P.J. (1982). Use of Congo red-polysaccharide interactions in enumeration of cellulolytic bacteria from the bovine rumen. *Appl. Environ. Microbiol.*, 43:777–780.

Thompson, J.D. and Gibson, T.J. and Plewniak, F. and Jeanmougin, F. and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 24:4876–4882.

Thomson, J.A. (1993). Molecular biology of xylan degradation. *FEMS Microbiol. Rev.*, 104:65–82.

Tomme, P. and Chauvaux, S. and Beguin, P. and Millet, J. and Aubert, J. and Claeyssens, M. (1991). Identification of a histidyl residue in the active center of endoglucanase D from *Clostridium thermocellum*. *J. Biol. Chem.*, 266:10313–10318.

Tomme, P. and van Beeumen, J. and Claeyssens, M. (1992). Modification of catalytically important carboxy residues in endoglucanase D from *Clostridium thermocellum*. *J. Biol. Chem.*, 267:4472–4478.

Tomme, P. and Warren, R.A.J. and Miller, R.C. and Kilburn, D.G. and Gilkes, N.R. (1995). Cellulose-binding domains - classification and properties. *Washington DC: American Chemical Society.*

Towbin, H. and Staehlin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci.*, 76:4350–4354.

Uitdehaag, J.C.M. and Mosi, R. and Kalk, K.H. and van der Veen, B.A. and Dijkhuizen, L. and Withers, S.G. and Dijkstra, B.W. (1999). X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nat. Struct. Biol.*, 6:432–436.

Wassenberg, D. and Liebel, W. and Jaenicke, R. (2000). Maltose-binding protein from the hyperthermophilic bacterium *Thermotoga maritima*: stability and binding properties. *J. Mol. Biol.*, 295:279–288.

Wisotzkey, J.D. and Jurtshuk, P. and fox, G.E. and Deinhard, G. and Poralla, K. (1992). Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for the creation of a new genus, *Alicyclobacillus gen. nov. Int. J. Syst. Bact.*, 42:263–269.

Withers, S.G. (2001). Mechanisms of glycosyl transferases and hydrolases. *Carbohydr. Polym.*, 44:325–337.

Wood, T.M. (1988). Preparation of crystalline, amorphous and dyed cellulase substrates. *Methods Enzymol.*, 160:19–25.

Yanish-Perron, C. and Vicira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33:103–119.

Zverlov, V.V. and Liebl, W. and Bachleitner, M. and Schwarz, W.H. (1998). Nucleotide sequence of *arfB* of *Clostridium stercorarium*, and prediction of catalytic residues of α-L-arabinofuranosidases based on local similarity with several families of glycosyl hydrolases. *FEMS Microbiol. Lett.*, 164:337–343.

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor Erwin Schneider for continuous support during all the ups and downs of this work. Especially for his willingness to discuss results (or no results) at all times with energy. Also thank you for a constant supply of hints and suggestions.

Thank you also to the following people who put time and energy into my PhD project in various ways:

Frank Scheffel for sharing a lab with me without going crazy and digging out his experiments on *A. acidocaldarius* from one of fifty folders when asked. His readiness to point out the funny side of zero transformants per g plasmid, vanishing proteins, turned off shakers and literature seminars was often the thing that stopped me from a change in profession[1].

Anja for sharing a lab with me during the earlier stages of my work. It was fun to have someone working at the opposite bench who had always seen the latest films, shows etc. Thanks also for a lot of help and discussions on *A. acidocaldarius*.

Leila Lo Leggio at the Centre for Crystallographic Studies (CCS), University of Copenhagen in Denmark for introducing me to protein crystallization. Her enthusiasm and fun when fiddling with crystals, atoms, computers and synchrotrons (or all of them at once) made my time in Denmark memorable. Thank you also for getting generally interested in CelA and its substrate specificities and for battling on with CelA when I was already safely back in Berlin.

Sine Larsen for having me at CCS and making my stay in the group a pleasant one.

Renu Kadirvelraj[2] also for her interest in CelA crystals, but mainly for nights at the opera, outings to Greater Copenhagen and Indian cooking.

The rest of the happy bunch at CCS for being helpful at all times, giving

[1] a vegetable store often seemed a good idea

[2] there aren't many Renus around, but adding the surname lends a more outlandish touch to the thesis

up space and computers, and the ingenious invention of 'Friday beer'.

Jens Laursen at the University of Copenhagen for X-ray fluorescence analysis of CelA.

Roland Schmid (University of Osnabrck) for performing N-terminal sequence analyses and getting the best out of what little protein we had.

Evert Bakker (University of Osnabrck) for sharing data on *A. acidocaldarius* and its enzymes.

Heidi Landmesser for helping with much of the lab equipment and for giving valuable advice on many lab procedures. Thank you also for organizing birthday breakfasts. They will be missed.

Birgit Sattler for lab work on purification of CelA.

Gabriele Brune for lab work on purification and solubilization of CelB.

Sabine for help in the lab and giving advice on many lab and lab-paperwork problems.

Betty for introducing me to the RZ85 and pointing out that the only proper thing to do on a Friday afternoon is to go canoeing[3].

Claudia for often discussing my project, but also for coming up with the more interesting topics during lunch break.

Anke for adding to the interesting topics.

Vivien for having an ever so slight readiness to laugh at microbes and their tormentors even in the bleakest of moments. And for making sure I crawled in the right direction after going out.

Dirk for knowing where things are in R224.

Sybille for parties on the outskirts of Berlin and for telling me at regular intervals to finish my PhD.

Oliwia for helping out with stuff from the next door lab and her appreciation of the 'Derya'.

Kirill for being genuinely interested in my project and giving me food for thought.

Robert Arndt for helping quickly with the computers even though there were a lot to take care of.

Ed Schwarz for solving many of the computer network problems in our building.

[3] preferably till the following Friday

Christian and Udo for helping with the computers when all else failed and especially Udo for discussing my project and giving his opinion.

My parents for their support and encouragement during the past years. Thank you for living in a place where I was sure to 'get away from it all' whilst still supplying the comforts of modern civilization.

This work was supported by fellowships of the Deutsche Bundesstiftung Umwelt (DBU), the European Molecular Biology Organization (EMBO), and the Federation of European Microbiology Societies (FEMS).

A Appendix

A.1 Curriculum vitae

18.12.71	Born in Farnworth, England
1975–79	Egerton County Primary School, England
1979–83	Jrgen Schule Flensburg (primary school), Germany
1983–91	Goethe-Gymnasium Flensburg (high school)
Oct. 1991 – Oct. 1993	Study of Biology at the Christian-Albrechts-Universitt zu Kiel
Jan. 1993 – June 1993	Work at the 'Forschungsstelle fr kosystem-forschung' (ecosystems research), Kiel
24.9.93	Vordiplom' in Biology
Oct. 1993 – June 1994	Study of Biology at the University of Edinburgh (Scotland)
Oct. 1994 – Jan. 1998	Study of Biology at the Humboldt-Universitt zu Berlin (HUB)
Oct. 1994 – Sept. 1995	Work in the laboratory of bacterial genetics (Prof. R. Borriss) at the HUB
June 1996 – Dec. 1996	Work at the Max Planck Institute of Molecular Plant Physiology, Golm (Dr. Fissahn)

Jan. 1998	Final degree passed with very good' distinction. Majors in microbiology, genetics and biochemistry, diploma thesis: Biochemical characterization of a xylan-degrading enzyme activity from *Alicyclobacillus acidocaldarius* ATCC27009'
Feb. 1998 – May 1998	Work in the laboratory of bacterial physiology (Prof. E. Schneider) at the HUB
Since 1st June 1998	PhD in biology in the laboratory of bacterial physiology at the HUB. PhD thesis: Cloning and Characterization of two glycosidases from the acidothermophile *Alicyclobacillus acidocaldarius* ATCC27009'
Oct. 2000 – Jan. 2001, Sept. 2001 – Oct. 2001	Crystallization of CelA during two stays at the 'Centre for Crystallographic Studies, Chemical Institute, University of Copenhagen' in Denmark (Dr. L. Lo Leggio, Dr. S. Larsen) during my PhD work.

.....
Kelvin Eckert

Berlin,

A.2 Publications

Eckert, K. and Schneider, E. (2003). A thermoacidophilic endoglucanase (CelB) from *Alicyclobacillus acidocaldarius* displays a high sequence similarity to arabinofuranosidases belonging to family 51 of glycoside hydrolases. *Eur. J. Biochem.* **270**: 3593-3602.

Eckert, K., Ernst, H.A., Schneider, E., Larsen, S. and Lo Leggio, L. (2003). Crystallization and preliminary X-ray analysis of *Alicyclobacillus acidocaldarius* endoglucanase CelA. *Acta Crystallogr. D. Biol. Crystallogr.* **59**: 139-141.

Eckert, K., Zielinski, F., Lo Leggio, L. and Schneider, E. (2002). Gene cloning, sequencing, and characterization of a familiy 9 endoglucanase (CelA) with an unusual pattern of activity from the thermoacidophile *Alicyclobacillus acidocaldarius* ATCC27009. *Appl. Microbiol. Biotechnol.* **60**: 428-436.

Lo Leggio, L., Kalogiannis, S., Eckert, K., Teixeira, S.C.M., Bhat, M.K., Andrei, C., Pickersgill, R.W. and Larsen, S. (2001). Substrate specificity and subsite mobility in *T. aurantiacus* xylanase 10A. *FEBS Letters* **509**: 303-308.

Talks

VAAM (Vereinigung fr Allgemeine und Angewandte Mikrobiologie), Jahrestagung 2003, short talk:

Eckert, K., Zielinski, F., Lo Leggio, L. and Schneider, E. (2003). Cloning and characterization of two glycosidases from the acidothermophile *Alicyclobacillus acidocaldarius*. *Biospektrum* Sonderausgabe der Jahrestagung der VAAM 2003 p. 51 (KE001).

Posters

Eckert, K. and Schneider, E. (1999). Characterisation of xylanase activity from *Alicyclobacillus acidocaldarius*. *Biospektrum* Sonderausgabe der Jahrestagung der VAAM 1999 p. 129 (P-M 24).

Eckert, K. and Schneider, E. (2000). Characterisation of a 103 kDa xylanase from *Alicyclobacillus acidocaldarius*. *Biospektrum* Sonderausgabe zum Kongress der DGHM, GHMP und VAAM 2000 p. 89 (13.P.9.14)

.....
Kelvin Eckert

Berlin,

A.3 Selbstndigkeitserklrung

Hiermit erklre ich, die vorliegende Arbeit selbstndig ohne fremde Hilfe verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet zu haben.

.....
Kelvin Eckert

Berlin,